

UNIVERSIDAD AUTÓNOMA DE MADRID

ESCUELA POLITÉCNICA SUPERIOR



Grado en Ingeniería en Informática

TRABAJO FIN DE GRADO

HERRAMIENTA DE ENTRENAMIENTO VOCAL

Autor: Jorge Casals de la Moneda
Tutor: Dr. Alberto Suárez González

julio 2018

HERRAMIENTA DE ENTRENAMIENTO VOCAL

Autor: Jorge Casals de la Moneda
Tutor: Dr. Alberto Suárez González

Dpto. Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
julio 2018

Resumen

Resumen

Saber proyectar la voz es un elemento importante para toda persona, sobre todo para aquellas que la usan como instrumento de trabajo, ya que permite que la voz pueda oírse de forma clara, que sea fácil de entender las palabras, que el sonido natural de la voz no se vea alterado en ningún sentido, tanto a volumen bajo como a volumen alto, manteniendo una producción vocal sana, libre de tensiones y sin lesionar los órganos de fonación.

El objetivo de este trabajo es desarrollar una aplicación que ayude al usuario a mejorar la proyección de su voz a lo largo del tiempo, obteniendo feedback de ella.

A partir del análisis de numerosos estudios consultados acerca de la proyección de la voz, hemos decidido implementar algoritmos como el Espectro y el LTAS (Espectro Medio a Largo Plazo) que han demostrado ser útiles para el análisis de la misma y nos permite determinar su grado de proyección una vez grabada por el propio usuario.

Para el desarrollo de esta aplicación usamos Python con Pycharm, Git y un programa llamado PRAAT que nos sirvió de guía para desarrollar el LTAS.

El resultado final ha sido el desarrollo de una aplicación que pudiera utilizarse como herramienta para proveer información acerca de la proyección de la voz en cada sesión de entrenamiento, permitiéndole al usuario proyectar mejor, con el máximo rendimiento y mínimo esfuerzo.

El trabajo ofrece un conjunto de consideraciones, conclusiones y recomendaciones que pudieran ser usados para perfeccionar la propuesta de aplicación.

Palabras Clave

Espectro, Espectro Medio a Largo Plazo, LTAS, Voz Humana, Proyección de la voz, Python, Pycharm, Git, PRAAT

Abstract

Knowing how to project the voice is an important element for every person, especially for those who use it as an instrument of work, since it allows the voice to be heard clearly, it is easy to understand the words, the natural sound of the voice is not altered in any sense, both at low volume and high volume, maintaining a healthy vocal production, free of tension and without damaging the organs of phonation.

The objective of this work is to develop an application that helps the user to improve the projection of his voice over time, obtaining feedback from it.

From the analysis of numerous studies consulted about the projection of the voice, we have decided to implement algorithms such as the Spectrum and the LTAS (Long Term Medium Spectrum) that have proved useful for the analysis of the same and allows us to determine its degree of projection once recorded by the user.

For the development of this application we use Python with Pycharm, Git and a program called PRAAT that served as a guide to develop the LTAS.

The final result has been the development of an application that could be used as a tool to provide information about the projection of the voice in each training session, allowing the user to project better, with maximum performance and minimum effort.

The work offers a set of considerations, conclusions and recommendations that could be used to refine the application proposal.

Palabras Clave

Spectrum, Long-term Average Spectrum, LTAS, Human Voice, Voice Projection, Python, Pycharm, Git, PRAAT

Agradecimientos

Soli Deo Gloria

Los siguientes agradecimientos están escritos en orden cronológico y no en orden de importancia necesariamente.

A mi tutor Dr. Alberto Suárez por tus ánimos, tu ayuda y confianza en mí. Gracias por todo lo que me has enseñado.

A mi madre: Por ayudarme en todo.

A mi esposa: Por velar por mí, ayudarme y darme fuerzas.

A Raquel y Atanasio: Por darme estabilidad y ejemplo en estos años.

A mi amigo Adoniram: Por ser mi consejero.

A mi padre Manuel: Por desarrollar mi interés por los juegos y darme ejemplo de buen trabajador.

A mi padre Jorge: Por darme ejemplo de estudio en la edad adulta.

A mi abuela Ardenia: Por desarrollar mi pasión por los misterios. RIP.

A mi tío Pepe: Por enseñarme a sonreír.

A mi abuela María: Por desarrollar mi pasión por la aventura.

A mi maestra de primaria Honoris Savón: Por salvarme del fracaso escolar.

Y a todos: los que como Scot Musser y Andy Venet se han preocupado por el trabajo y me han ayudado. A todos los que se preocupan llamándome y escribiendome a diario para saber cómo iba el proyecto. A Cristina, Luis Felipe, Noemí, Eliseo y María López, Carmen y Nico, Antonio, Alex, entre muchos otros que me han preguntado constantemente, mostrando su preocupación y deseo de que todo salga bien. Gracias.

A la Universidad Autónoma de Madrid por darme la oportunidad de estudiar a destiempo y con contratiempos. A secretaría por ayudarme en todos mis papeles y gestiones, especialmente en medio de la enfermedad de mi esposa, aún cuando por faltas mías les lié más de la cuenta. A todos los profesores de la carrera, en especial una vez más a Alberto Suárez por sus clases de gran nivel.

Por último a la profesora Susana Holgado por darme la bienvenida a la UAM y confiar en mí.

Índice general

Índice de Figuras	IX
Índice de Tablas	XI
1. Introducción	1
1.1. Motivación del proyecto	1
1.2. Objetivos	1
1.3. Organización de la memoria	1
2. Proyección de la voz. Estado del arte	3
2.1. El sonido	3
2.1.1. El tracto vocal	4
2.1.2. Análisis Espectral	5
2.1.3. Espectro Medio a Largo Plazo o LTAS	5
2.1.4. Los formantes	5
2.2. Proyección	6
2.2.1. La frecuencia mágica: 3-4 kHz	7
2.2.2. El esfuerzo vocal: 5-8 kHz	8
3. Cálculo de la Proyección	9
3.1. Algoritmo Global	9
3.2. LTAS	10
3.3. Área bajo la curva	11
4. Sistema, diseño y desarrollo	13
4.1. Lenguaje de programación, entorno y herramientas elegidas	13
4.2. Requisitos: Historias de usuario	13
4.3. Diseño de clases	15
4.4. Diagrama de flujo de la aplicación	21
4.5. Descripción de la aplicación desarrollada	22
4.6. Pruebas y Resultados	26

4.6.1. Test de los algoritmos Espectro y LTAS	26
4.6.2. Test finales de los algoritmos y la proyección	28
4.7. Encuesta de satisfacción	30
4.7.1. Preguntas de la encuesta	30
4.7.2. Participantes de la encuesta	30
4.7.3. Resultados de la encuesta	31
5. Conclusiones y trabajo futuro	33
5.1. Conclusiones	33
5.2. Trabajo Futuro	33
Glosario de acrónimos	35
Bibliografía	36
A. ANEXO: Encuesta de la aplicación	39

Índice de Figuras

2.1. Armónicos de una onda periódica	4
2.2. Tracto vocal	4
2.3. Ondas de la voz y su espectro de frecuencia	5
2.4. De izquierda a derecha: El espectro y su ltas	5
2.5. F1 vs F2 al pronunciar las vocales en español	6
2.6. Anatomía del oído	7
2.7. Curva de Fletcher y Munson	7
2.8. Espectro del llanto de neonatos al sufrir dolor	8
3.1. LTAS: Área de proyección	10
3.2. Single Transformation LTAS	10
3.3. Short-term averaging LTAS	11
4.1. Diagrama de clases. Controlador Login.	15
4.2. Diagrama de clases: Controlador Register	16
4.3. Diagrama de clases: Controlador trainer	16
4.4. Diagrama de clases: Controlador trainings	17
4.5. Diagrama de clases: Servicio de usuario	17
4.6. Diagrama de clases: Servicio de Entrenamiento	17
4.7. Diagrama de clases: Servicios de Espectro	18
4.8. Diagrama de clases: Servicios de LTAS	18
4.9. Diagrama de clases: Entidad Espectro	18
4.10. Diagrama de clases: Entidad LTAS	18
4.11. Diagrama de clase: Entidad Audio	19
4.12. Diagrama de clases. Entidad Entrenamiento	19
4.13. Diagrama de clases: Entidad Lectura	19
4.14. Diagrama de clases: Entidad Usuario	19
4.15. Diagrama de clases: Algoritmo Creador de LTAS	19
4.16. Diagrama de clases: Algoritmo Creador de Espectro	20
4.17. Diagrama de clases: Algoritmo de sonido	20

4.18. Diagrama de clases: Algoritmo calculador de proyección	20
4.19. Diagrama de clases: Repositorio de Usuario	21
4.20. Diagrama de clases: Repositorios de Ltas y Espectro	21
4.21. Diagrama de flujo de la aplicación	22
4.22. Aplicación: Ventana Login	22
4.23. Aplicación: Ventana Entrenamientos Previos	23
4.24. Aplicación: Comenzar nuevo entrenamiento	23
4.25. Aplicación: Ventana Sesión de Entrenamiento	24
4.26. Aplicación: Grabando la lectura	24
4.27. Aplicación: Grabación terminada	25
4.28. Aplicación: Entrenamiento añadido	26
4.29. Comparativa LTAS: nuestra	27
4.30. Comparativa LTAS: del PRAAT	27
4.31. Comparativa de Espectro de Frecuencias	28
4.32. LTAS del test 1	29
4.33. LTAS del test 2	29
4.34. LTAS del test 3	29
4.35. LTAS del test 4	29
4.36. LTAS del test 5	29
4.37. LTAS del test proyectando	30
4.38. LTAS del test no proyectando	30

Índice de Tablas

4.1. Resultados de 5 entrenamientos con igual proyección	28
4.2. Resultados de 2 tests: Proyectando y no proyectando	30

1

Introducción

1.1. Motivación del proyecto

La voz es el instrumento fundamental para comunicarnos con los demás y en el caso de los profesores y comunicadores es también su herramienta de trabajo.

El uso correcto de la voz implica un aprendizaje. Es necesario conocer el mecanismo de la fonación y los recursos de la voz, para tomar conciencia de los propios hábitos, para corregirlos mediante el ejercicio y la automatización de los mecanismos más adecuados.

La proyección de la voz ha sido y es objeto de estudio desde hace mucho tiempo. Los profesores de canto por ejemplo, de forma empírica le indicaban a sus alumnos a adoptar ciertas posturas y sentir ciertas vibraciones para conseguir la proyección deseada. El oído del profesor que estaba lo suficientemente entrenado, le daba feedback para que el alumno pudiera seguir en ese camino al proyectar su voz.

En el caso del habla, en la actualidad es difícil obtener información acerca de la calidad de la proyección de la voz sobre todo en aquellas personas que necesitan o deseen mejorar este aspecto. Esto nos motivó a realizar e implementar una aplicación que les ayudara a desarrollar habilidades para proyectarla mejor, lo que permitiría forzar menos las cuerdas vocales e intentar protegerlas de los daños del uso continuado, especialmente a profesores, oradores, comentaristas, locutores, abogados, actores de teatro, entre otros.

1.2. Objetivos

Nuestro principal objetivo es desarrollar una aplicación que dé información al usuario de la proyección de su voz y de su evolución en el tiempo. Para lograr desarrollar esta aplicación primero es necesario comprender la proyección de la voz humana e investigar su detección, para posteriormente desarrollar algoritmos que detecten la proyección de la voz que serán usados en el desarrollo de dicha aplicación.

1.3. Organización de la memoria

El trabajo quedó estructurado de la siguiente manera:

En el capítulo Estado del Arte exponemos definiciones básicas acerca del sonido, la voz humana, análisis espectral y formantes entre otros, finalizando con una aproximación a la proyección de la voz, en la cual intervienen todos los conceptos antes expresados.

En el siguiente capítulo, al que denominamos Algoritmos, realizamos una descripción de los algoritmos utilizados para detectar la proyección de la voz, en especial el LTAS o espectro medio a largo plazo.

Sistema Desarrollado es el tercer capítulo, en el se describen las historias de usuario, el diseño y los detalles de implementación y pruebas de la herramienta de entrenamiento vocal. También se incluye una encuesta de satisfacción.

Las Conclusiones, donde se sintetiza los logros fundamentales alcanzados y algunas recomendaciones que orientan los aspectos en los que a juicio del autor se debe proyectar al futuro la continuación de este trabajo.

Contiene además una Bibliografía que soporta el abanico de autores y fuentes consultadas.

2

Proyección de la voz. Estado del arte

2.1. El sonido

Para comprender la proyección de la voz, es necesario en primer lugar explicar las propiedades básicas del sonido para luego, al hablar de estas en relación con la voz, sepamos a qué nos referimos.

Altura o Tono: Sirve para identificar el sonido como agudo, medio o grave. Son la consecuencia de la frecuencia de las ondas sonoras. A mayor frecuencia más agudo se percibirá el sonido. El rango de audición humano está comprendido entre los 20 Hz y los 20kHz. Este es el rango de frecuencias que escogimos para el Espectro de Frecuencias, calculado mediante la Transformada de Fourier Discreta.

Intensidad: Es la cantidad de energía acústica que contiene un sonido, es decir, lo fuerte o suave de un sonido. La intensidad viene determinada por la potencia, que a su vez está determinada por la amplitud y nos permite distinguir si el sonido es fuerte o débil.

Timbre o Color: Los sonidos se pueden distinguir aunque tengan la misma frecuencia e intensidad por el timbre o color de la voz. Por ejemplo, cuando hablan dos personas distintas con el mismo tono e intensidad, podríamos diferenciarlas por el timbre de la voz de cada una. Un ejemplo más evidente, es la diferencia de sonido entre mujeres, hombres y niños, suponiendo que hablaran en el mismo tono e intensidad. Esta es quizás la característica de la voz más subjetiva, aunque como veremos más adelante, está relacionada íntimamente con los formantes(concepto que veremos más adelante).

Ondas periódicas y armónicos: El sonido de la voz está formado por ondas periódicas que tienen varias frecuencias, por lo que cualquier sonido de una onda periódica, podrá ser descompuesto como una superposición de la frecuencia fundamental y armónicos superiores. La frecuencia fundamental es la frecuencia más baja de una forma de onda periódica. Los armónicos superiores son vibraciones de frecuencias más altas que son múltiplos enteros de la frecuencia fundamental.

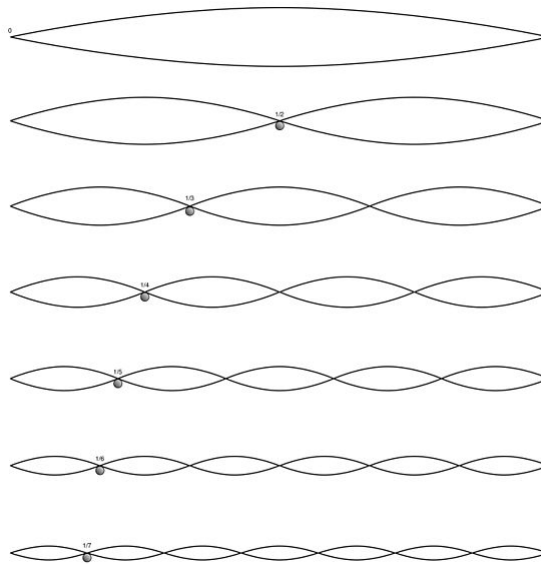


Figura 2.1: Armónicos de una onda periódica

Fuente: <https://commons.wikimedia.org/wiki/File:Overtone.jpg>

2.1.1. El tracto vocal

En la emisión de la voz participan los pulmones que aportan la energía necesaria para hacer vibrar las cuerdas vocales entre sí y emitir lo que sería la frecuencia fundamental del sonido de la voz. A partir de ahí, las ondas del sonido recorren el camino del tracto vocal hacia la boca. La anatomía del tracto vocal potencia algunos sonidos a través de su resonancia.

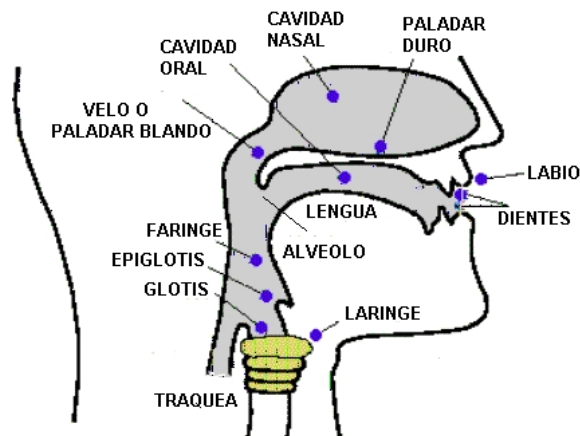


Figura 2.2: Tracto vocal

Fuente: <https://es.wikipedia.org/wiki/Archivo:Fig21.jpg>

El tracto vocal tiene como función principal articular y formar los sonidos que luego serán escuchados por otras personas para su interpretación, por ejemplo, la pronunciación de las vocales o consonantes. Al conjunto de órganos que lo forman, se le asocian todas las propiedades conocidas y aún por conocer de la proyección.

2.1.2. Análisis Espectral

Es el proceso mediante el cual se cuantifica la intensidad de cada frecuencia de una fuente de ondas sonoras. El concepto se aplica a todos los fenómenos ondulatorios (luz, sonido o electromagnetismo). En nuestro análisis, partimos de un archivo de sonido y creamos un espectro de frecuencias mediante la Transformada de Fourier Discreta, asumiendo que el sonido es expresado en pascales (Pa) y el espectro es expresado en Pa/Hz. Este cálculo en la aplicación que proponemos será expuesto más adelante.

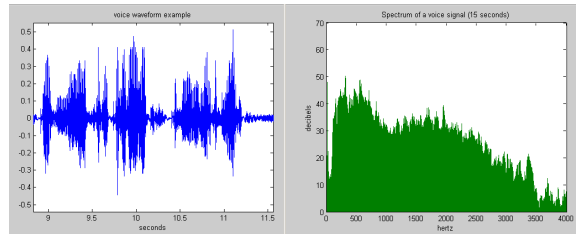


Figura 2.3: Ondas de la voz y su espectro de frecuencia

Fuente: https://commons.wikimedia.org/wiki/File:Voice_waveform_and_spectrum.png

2.1.3. Espectro Medio a Largo Plazo o LTAS

El espectro medio a largo plazo o LTAS (del inglés: Long-term Average Spectrum) está formado por las medias de ventanas de análisis cortas en un espectro. En nuestro caso, la ventana de análisis es de unos 100 Hz, por ser el valor que nos ha aportado mayor definición. El uso del LTAS ha sido valorado por muchos autores cómo el método más efectivo hasta la fecha para analizar las características de la voz [1]. En el capítulo Cálculo de la Proyección hablaremos más del LTAS.

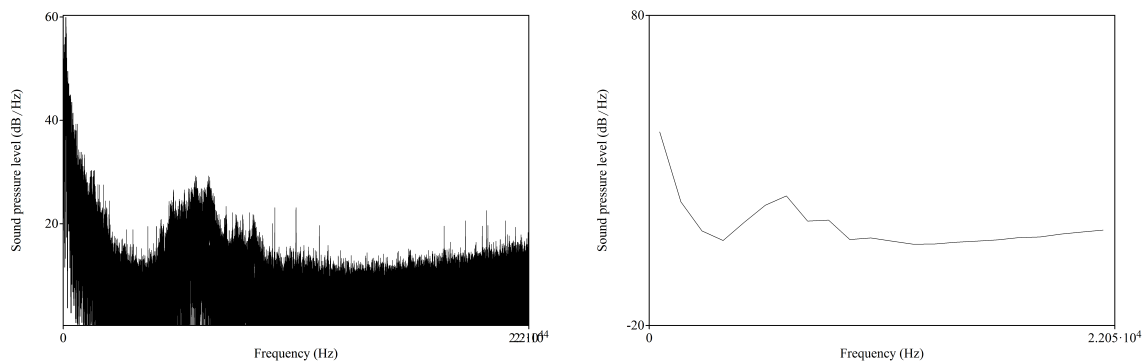


Figura 2.4: De izquierda a derecha: El espectro y su ltas

2.1.4. Los formantes

Se define un formante como el armónico de una nota que es aumentada por una resonancia [2]. En el tracto vocal humano se producen resonancias al emitir sonidos, que son denominadas formantes. Debido a esto, en fonética un formante puede usarse para referirse a una resonancia o al máximo espectral que una resonancia produce [3]. Los investigadores y analistas de la fonética relacionan los formantes con los picos de amplitud en el espectro de frecuencias del sonido, siendo estos picos una estimación de las resonancias del tracto vocal.

Origen de los formantes

En el tracto vocal se produce la energía de los formantes. La vibración periódica de las cuerdas vocales produce una serie de armónicos. Los armónicos cuya frecuencia esté cerca de una resonancia del tracto vocal, son potenciados y pasan libremente a través de este, produciendo un formante. Aquellos armónicos cuyas frecuencias no están cerca de las frecuencias de una resonancia se debilitan formando valles entre los picos de cada formante [4]. En el análisis de la voz los formantes más importantes y a menudo los únicos que se pueden detectar son: F1, F2, F3, F4, F5. Los primeros tres formantes (F1-F3) se relacionan a menudo con la pronunciación de las palabras y los últimos formantes (F4 y F5) con la calidad de la voz.

El proceso de articulación determina las frecuencias de los formantes vocales. Se ha identificado las partes de la anatomía vocal que se asocia con las frecuencias de los formantes. La apertura de la mandíbula, que constriñe el tracto vocal hacia el final de la glotis y lo expande en el extremo del labio, es el factor decisivo para F1. La frecuencia de este formante aumenta a medida que la mandíbula se abre más. El segundo formante F2 es más sensible a la forma del cuerpo de la lengua, y el tercer formante F3 a la punta de la lengua [5].

Aunque todos los formantes dependen de la anatomía del tracto vocal, especialmente los dos últimos (F4 y F5) se pueden modificar usando sus músculos, y en gran medida son los que influyen en la calidad de la proyección [1][6].

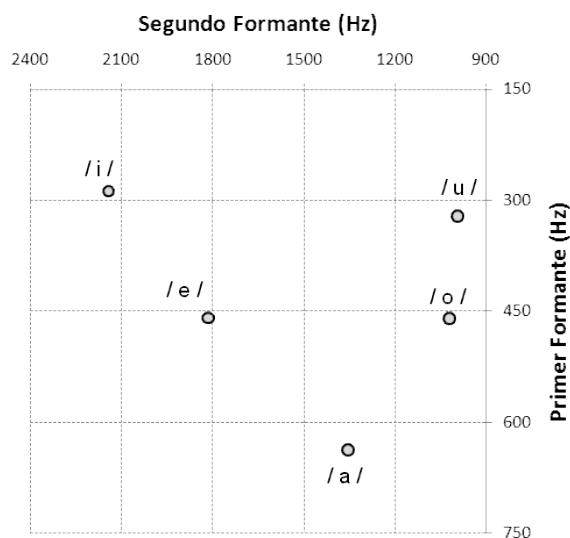


Figura 2.5: F1 vs F2 al pronunciar las vocales en español

Fuente: https://commons.wikimedia.org/wiki/File:Spanish_Vowel_Formants_Bradlow1995.png

2.2. Proyección

La proyección de la voz ha sido objeto de estudio a lo largo de los años, sin embargo, aún se cree que hay variables desconocidas que intervienen en la proyección de la voz, por ejemplo, la impedancia mecánica de la pared del tracto vocal observada en modelos 3D realizados a partir de resonancias magnéticas del tracto vocal [7]. Es importante aclarar que este trabajo no pretende definir la proyección en toda su magnitud, teniendo en cuenta que aún en la actualidad está lejos de ser completada, pero si usar lo que se sabe de ella para detectarla.

La proyección de la voz se puede ejemplificar de la manera siguiente: una persona habla en dos ocasiones, en una proyecta y en la otra no, los dos sonidos emitidos tienen la misma intensidad

desde las cuerdas vocales, no obstante, el sonido que se proyecta una vez que atraviesa el tracto vocal y sale de la boca tiene una intensidad mayor que el no proyectado.

2.2.1. La frecuencia mágica: 3-4 kHz

Receptor: El oído humano

El oído humano posee una especial sensibilidad a las frecuencias de 3-4 kHz. El canal auditivo es parte del oído externo, y amplifica el sonido de tal manera que puede ser convertido a señales electroquímicas por el oído medio e interno. Podemos oír mejor en el rango 3-4 kHz porque la frecuencia resonante del canal auditivo está en ese rango.

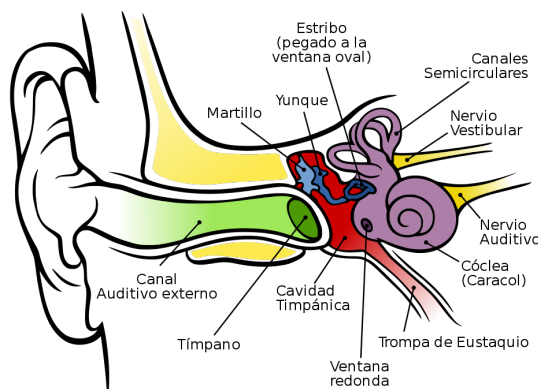


Figura 2.6: Anatomía del oído

Fuente: https://commons.wikimedia.org/wiki/File:Anatomy_of_the_Human_Ear_es.svg

El oído humano no escucha todas las frecuencias con la misma intensidad. En la década de los 30, los investigadores Fletcher y Munson fueron los primeros en medir y publicar un conjunto de curvas que mostraban la sensibilidad del oído humano a la amplitud del sonido comparada con la frecuencia. Las curvas mostraron que el oído es más sensible a los sonidos en el rango entre 3,000-4,000 Hz. Las frecuencias fuera de este rango deben ser emitidas más alto, si se quiere sean escuchadas con la misma precisión.

El ser humano escucha frecuencias alrededor de los 3,500 Hz de un sonido unos 10 a 20 dB más alto que otras frecuencias. Esta característica del oído humano es la responsable de que en las óperas se pueda escuchar a los cantantes por encima de la orquesta, ya que ellos aumentan la amplitud de su voz en esta zona del espectro donde el oído del público es especialmente sensible.

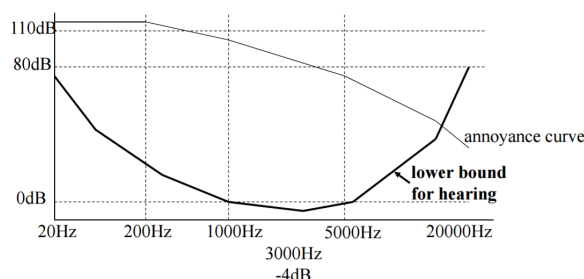


Figura 2.7: Curva de Fletcher y Munson

Fuente: https://commons.wikimedia.org/wiki/File:Fletcher-Munson_Curve.png

También se plantea, que pudiera ser un mecanismo de supervivencia que permite resaltar el sonido de los bebés cuando lloran. Por ejemplo, se han encontrado picos muy importantes en el llanto de neonatos sobre la zona 3.5 kHz cuando sufren dolor [8].

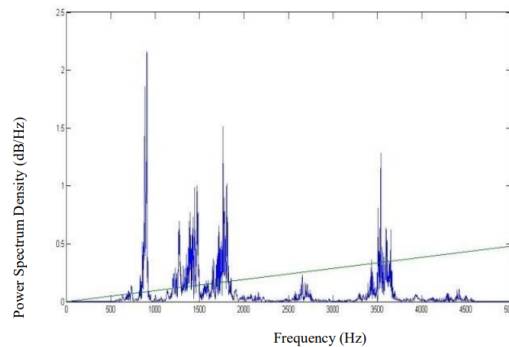


Figura 2.8: Espectro del llanto de neonatos al sufrir dolor

Fuente: [8]

Considerando este aspecto, sería conveniente mejorar la amplitud de nuestra voz en las frecuencias 3-4 kHz, pues si bien esta no mejora significativamente su amplitud (que sí lo hace), las personas podrían escucharnos más alto debido a que sus oídos distinguen mejor los sonidos en esta franja de frecuencias.

Emisor: Sinergia de formantes y el formante del orador

Se habla del formante del orador cuando la unión de los formantes F3 y F4 alrededor del área 3.5 kHz, produce una sinergia y aumenta de manera importante la amplitud del sonido emitido por la voz. Según algunos estudios el formante 4 se acerca al formante 3 en una vecindad de 3.5 kHz, provocando una sinergia que aumenta significativamente la intensidad de la voz [6]. Estos formantes se retroalimentan uno al otro en esta área, provocando que la potencia del sonido sea mayor al sumarlos que la suma de sus potencias por separado [6].

Concretamente cada vez que dos formantes se acercan a la mitad de su distancia original, las amplitudes de los formantes aumentan en 6 dB y el valle entre ellos en 12 dB. Debemos aclarar que no hay relación lineal entre el formante del orador (3.5 khz) y el tono de la voz, da igual si la voz del orador es grave o aguda. Todas las personas tienen el formante del orador en la misma frecuencia aproximadamente 3.5 khz [6][1].

2.2.2. El esfuerzo vocal: 5-8 kHz

Es interesante subrayar que en algunos estudios se sospecha que el esfuerzo de las cuerdas vocales está en la zona 5-8 kHz, ya que esta zona muestra un área bajo la curva del espectro proporcional al esfuerzo de las cuerdas vocales[9]. Esto no lo usamos, ya que hay resultados contrarios en otros estudios y se asocia también a la existencias de otros formantes(a partir de F5). Pero convendría seguirle la pista a los últimos estudios sobre esta zona para de ser confirmada la teoría, mejorar la aplicación.

3

Cálculo de la Proyección

3.1. Algoritmo Global

Las características que intervienen en la proyección de la voz son[6]:

1. Un pico en 3.5 kHz
2. La diferencia de intensidad 3-4 kHz con otras bandas
3. La frecuencia de F4
4. La diferencia de frecuencias entre F3 y F4
5. La intensidad del sonido
6. Área de F1

Nuestro algoritmo cubre las primeras 4 características, las dos restantes (como ya se dijo acerca de la proyección en estado del arte) dependen de la configuración fija de la anatomía del tracto vocal.

Para calcular el nivel de proyección de un sonido, transformamos la señal acústica en Espectro de frecuencias. A partir del espectro de frecuencias calculamos el LTAS. Luego hallamos el área bajo la curva entre las frecuencias 3 kHz y 4 kHz, siendo esta el área de proyección, y también hallamos el área bajo la curva entre las frecuencias 2 kHz y 5 kHz, siendo esta el área que rodea al área de proyección. El resultado final, es una proporción entre las dos áreas. El cual por supuesto está entre 0 y 1, ya que el área entre 3 kHz y 4 kHz está contenida en el área entre 2 kHz y 5 kHz. Este último valor es multiplicado por 10 para hacer que el valor esté siempre entre 1 y 10.

Haciendo esto incluimos en la medición el pico en 3.5 kHz, la diferencia de intensidad 3-4 kHz con la banda que la rodea 2 kHz – 5 kHz. Y por supuesto también incluimos la posible sinergia entre F3 y F4 dentro de 3 kHz – 4 kHz. El área de F4 que caiga por encima de 4 kHz influye negativamente pues mientras más lejos esté F4 de F3, menos proyección habrá, según hemos visto en el capítulo Estado del Arte.

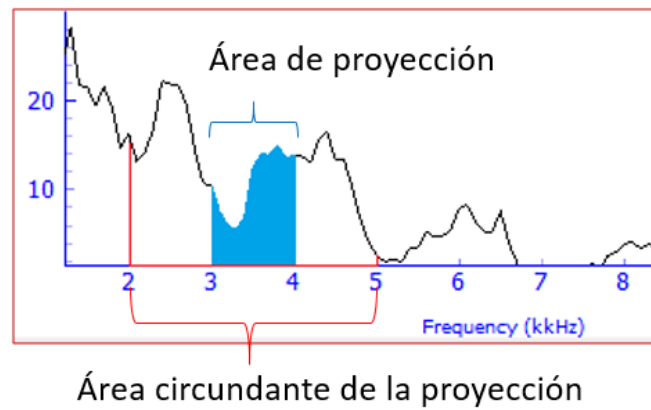


Figura 3.1: LTAS: Área de proyección

3.2. LTAS

El LTAS significa Espectro Medio a Largo Plazo. Según expertos el LTAS tiene una relación física directa con la localización de las resonancias del tracto vocal[10]. El LTAS pertenece a la clase de métodos de estimación de densidad de intensidad espectral. Tiene dos tipos de implementación:(es importante decir que en todas las transformaciones de señal de audio a espectro de frecuencias se hace un relleno de ceros al final de la señal, hasta que el número de valores alcance una potencia de dos)

1. **Transformación 1 a 1 (Single Transformation):** Transformar una señal de audio a Espectro de frecuencias. Luego se promedian rangos de frecuencia de un tamaño fijo, normalmente 50 Hz - 100 Hz.

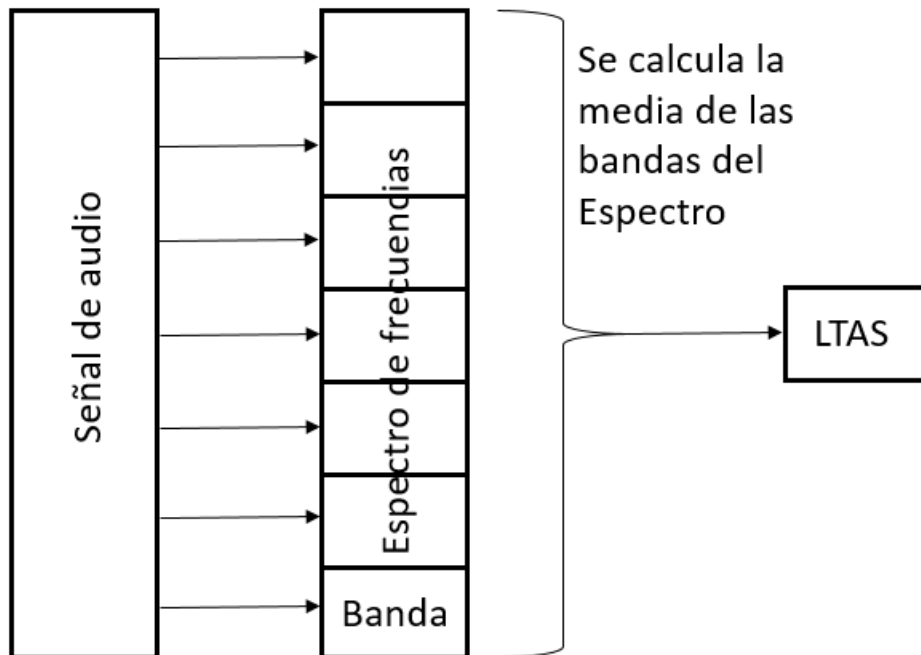


Figura 3.2: Single Transformation LTAS

2. **Media a corto plazo (Short-term averaging method):** También conocido como el método de Welch's, divide la señal en franjas superpuestas con las cuales hace pequeños espectros de frecuencia. Finaliza con la media de los espectros de frecuencia que se crearon de las franjas de la señal de audio.

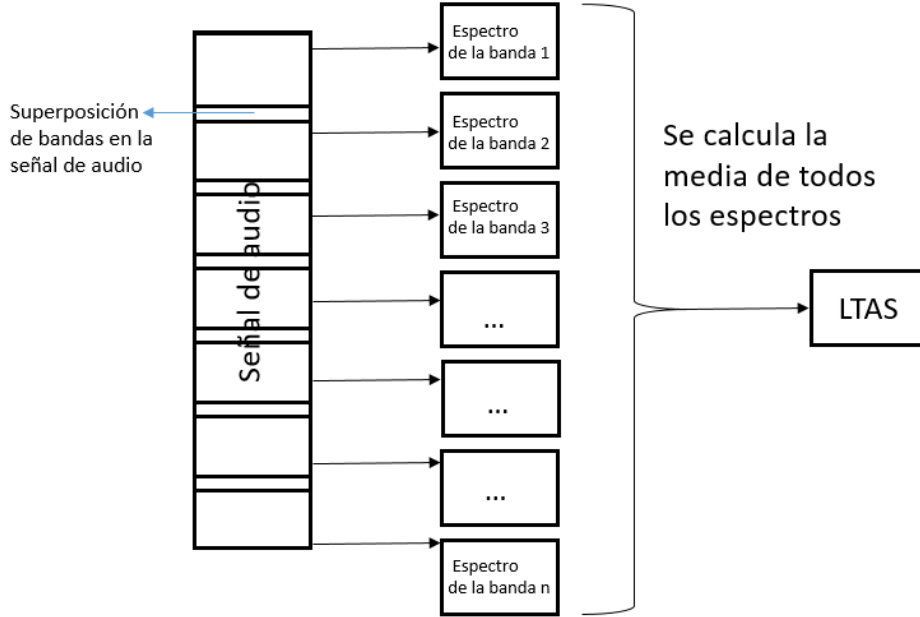


Figura 3.3: Short-term averaging LTAS

En nuestro trabajo implementamos el primer método, debido a que contábamos con la prestigiosa aplicación de Software Libre Praat para verificar nuestros valores. Y esta aplicación junto con el método 2, ha sido usado por los estudios en cuyos resultados basamos nuestro algoritmo. Para el método 2, el usado en esta aplicación, las bandas de frecuencias más usadas en estudios de la voz son 50 Hz [11] y 100 Hz. Nosotros optamos por hacer la media con una banda de 100 Hz [12].

Es interesante saber que algunos autores dicen que el LTAS se estabiliza después de 30 o 40 segundos de habla. Esta estabilización está relacionada con ruidos externos y/o exceso de pronunciación de algunas consonantes, que pudieran alterar el LTAS a muy corto plazo. Por ejemplo: 'P', 'B', 'S'. Esto pudiera usarse para mejoras futuras.

3.3. Área bajo la curva

El área bajo de una curva discreta la calculamos hallando el área de los trapecios que forman sus puntos.

Dado que el área de los trapecios se calcula con la siguiente fórmula:

$$A_t = (B_1 + B_2) * C/2$$

Dado que el área total de la curva es la suma de las áreas:

$$A_{total} = \sum_{n=1}^{220} A_n$$

Y que el valor C es constante, dado que en nuestro caso vale 100 Hz:

$$A_{total} = C/2 * (B_1 + 2 * (\sum_{n=2}^{220} B_n) + B_{221})$$

Lo que nos deja la siguiente fórmula empleada en nuestro programa:

$$A_{total} = C * ((B_1 + B_{221})/2 + (\sum_{n=2}^{220} B_n))$$

Con la anterior fórmula minimizamos los cálculos innecesarios del programa, mejorando el performance. Para terminar, dado que dividimos áreas calculadas sobre el mismo intervalo de frecuencias $C = 100Hz$, eliminamos la variable C de la ecuación anterior para mejorar aún más el performance cálculo.

4

Sistema, diseño y desarrollo

En este capítulo del trabajo explicaremos la aplicación desarrollada.

4.1. Lenguaje de programación, entorno y herramientas elegidas

Cuando iniciamos este proyecto, desconocíamos si era posible medir la proyección de la voz, por lo que decidimos adoptar una metodología ágil, en este caso una adaptación de Scrum, que nos permitiera si fuera necesario, cambiar el rumbo a lo largo del desarrollo del mismo.

Ante la no existencia de un equipo de múltiples desarrolladores adaptamos el framework de trabajo a nuestras necesidades. Tutor y autor nos reunimos con una frecuencia aproximada de 2 semanas, siendo esta la duración de nuestro Sprint. En esas reuniones se analizaron los logros alcanzados (Demostración), valorándose los resultados y las posibles mejoras a tener en cuenta (Retrospectiva), posteriormente se crearon nuevos objetivos para las siguientes dos semanas (Planificación).

Elegimos Python como lenguaje de programación orientado a objetos, no solo por ser libre, sino porque como lenguaje posee una flexibilidad y agilidad extraordinaria que nos permitió avanzar muy rápido en la implementación y enfocarnos más en la gran complejidad que ha supuesto entender y escribir los algoritmos antes expuestos. Para programar en este lenguaje usamos la guía de estilos y buenas prácticas de su creador.

Como framework de desarrollo hemos usado una licencia gratis para estudiantes de PyCharm de JetBrains, por ser de todos los entornos valorados, el que a nuestro juicio nos brindó mayor usabilidad y facilidad de uso. También hicimos uso de GIT como herramienta de control de versiones, teniendo todo el código en un repositorio público en GitHub.

Para la Transformada de Fourier Discreta y algunas funciones matemáticas usamos la librería Numpy.

4.2. Requisitos: Historias de usuario

En esta sección del capítulo detallaremos los requisitos del sistema.

Descripción de requisitos funcionales

1. Gestión de usuarios:

- **RF1:** El usuario entra en la aplicación y debe ser capaz de identificarse y acceder a todos sus datos y entrenamientos.
- **RF2:** En cualquier ventana de la aplicación el usuario debe ser capaz, de cerrar sesión sin salir de la aplicación y que otro usuario pueda entrar.

2. Gestión de entrenamientos:

- **RF3:** El usuario al entrar en la pantalla de entrenamientos, debe poder ver sus entrenamientos en una tabla.
- **RF4:** El usuario al entrar en la pantalla de entrenamientos y seleccionar un entrenamiento, debe ser capaz de escuchar su entrenamiento.
- **RF5:** El usuario al entrar en la pantalla de entrenamientos y seleccionar un entrenamiento, debe ser capaz de ver su Espectro.
- **RF6:** El usuario al entrar en la pantalla de entrenamientos y seleccionar un entrenamiento, debe ser capaz de ver su Espectro Medio a Largo Plazo.
- **RF7:** El usuario al entrar en la pantalla de entrenamientos debe ser capaz de ver el historial de los valores de su proyección.
- **RF8:** El usuario estando en la pantalla de entrenamientos debe ser capaz de comenzar un nuevo entrenamiento, lo cual le llevará a la pantalla de Nuevo entrenamiento.

3. Nuevo entrenamiento:

- **RF9:** El usuario estando en la pantalla de Nuevo entrenamiento, podrá comenzar un nuevo entrenamiento, seleccionando el texto de la lectura que va a realizar. A continuación el sistema deberá grabar la voz del usuario hasta que éste decida parar.
- **RF10:** El usuario podrá escuchar su grabación y decidir si la repite o no.
- **RF11:** El usuario podrá ver su puntuación de proyección al terminar de grabar.
- **RF12:** El usuario podrá dar por finalizado el entrenamiento y una vez hecho esto, no se podrá eliminar del sistema.

Descripción de los requisitos no funcionales

1. Interfaz y Usabilidad:

- **RNF1:** Textos de la aplicación en inglés para ser usada en cualquier país.
- **RNF2:** Interfaz sencilla e intuitiva para uso de cualquier persona.
- **RNF3:** Textos cortos y con fuente homogénea

2. Portabilidad:

- **RNF4:** Cómo mínimo deberá funcionar en Windows.
- **RNF5:** La aplicación deberá ser de fácil instalación.

3. Seguridad:

- **RNF6:** La aplicación debe admitir tener varios usuarios.

- **RNF7:** No es prioridad garantizar la seguridad de los datos del sistema, ya que no contiene información sensible ni personal.

4. Rendimiento:

- **RNF8:** El tiempo de respuesta no debe superar los 2 segundos.

4.3. Diseño de clases

La aplicación está compuesta principalmente por 6 módulos: Interfaz Gráfica, Controladores, Servicios, Entidades, Algoritmos y Repositorios. Un módulo está definido por el conjunto de clases que contiene.

- **Interfaz Gráfica:** Implementamos la interfaz gráfica de la aplicación usando PyQt, que es un binding de la biblioteca gráfica QT de C++ para el lenguaje de programación Python, disponible para Windows, Linux y Mac OS. La interfaz gráfica está compuesta por archivos tipo *.ui. Cada ventana o control visual está definido en este módulo, cuya responsabilidad es la presentación de la aplicación y no tiene conocimiento alguno del dominio o lógica de esta.
- **Controladores:** Los controladores son los encargados de dirigir el flujo de la aplicación. El usuario solicita una acción en la interfaz gráfica, donde hay un controlador que suscrito al evento de dicha acción, escucha dicha solicitud y orquesta los pasos que la componen mediante servicios y algoritmos, para posteriormente devolver una respuesta al usuario. Este conjunto de clases son solo orquestadores entre la UI y los servicios. Ejemplos de controladores son: Login, Register, Trainer, Trainings. También definimos una clase llamada ViewsNavigator que se encarga de navegar entre las vistas centralizando esta funcionalidad y una clase Session que contiene los datos de la sesión actual del usuario.

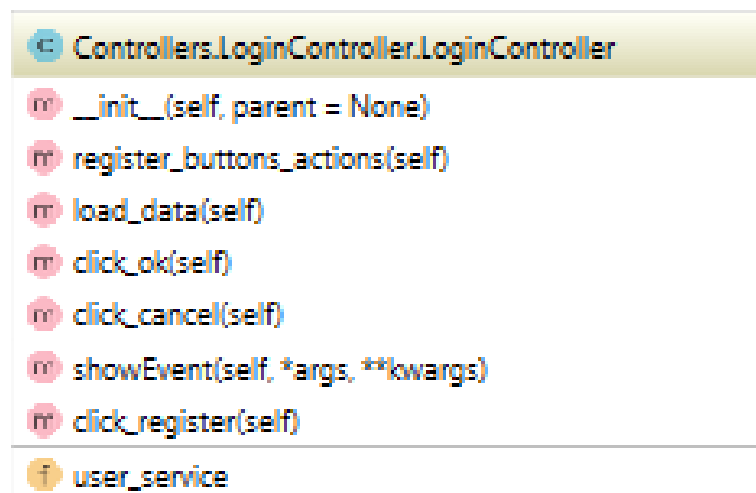


Figura 4.1: Diagrama de clases. Controlador Login.

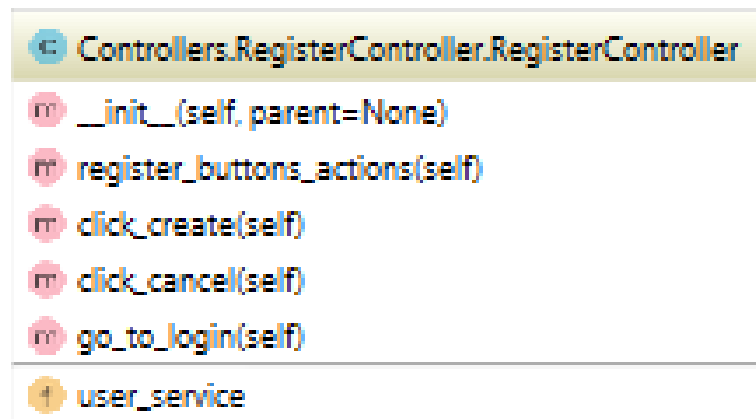


Figura 4.2: Diagrama de clases: Controlador Register

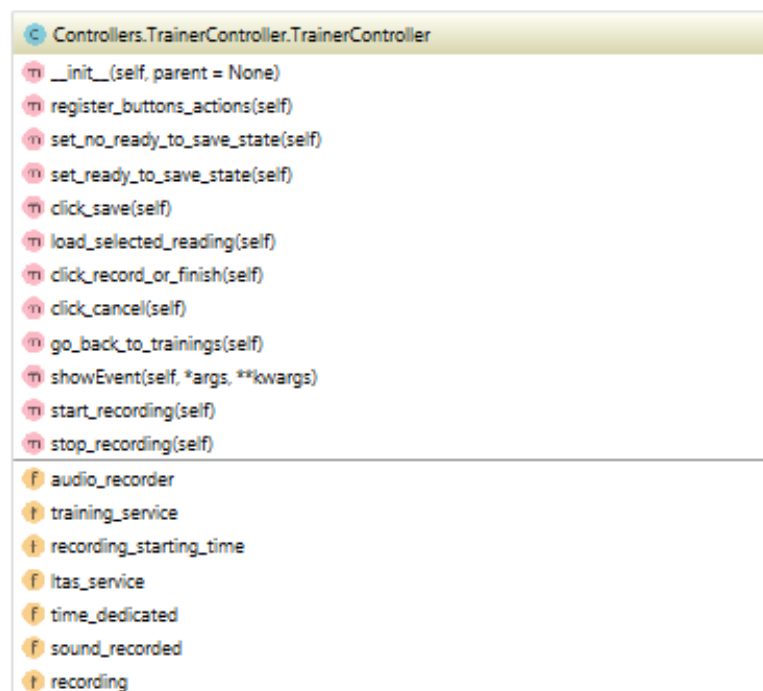


Figura 4.3: Diagrama de clases: Controlador trainer

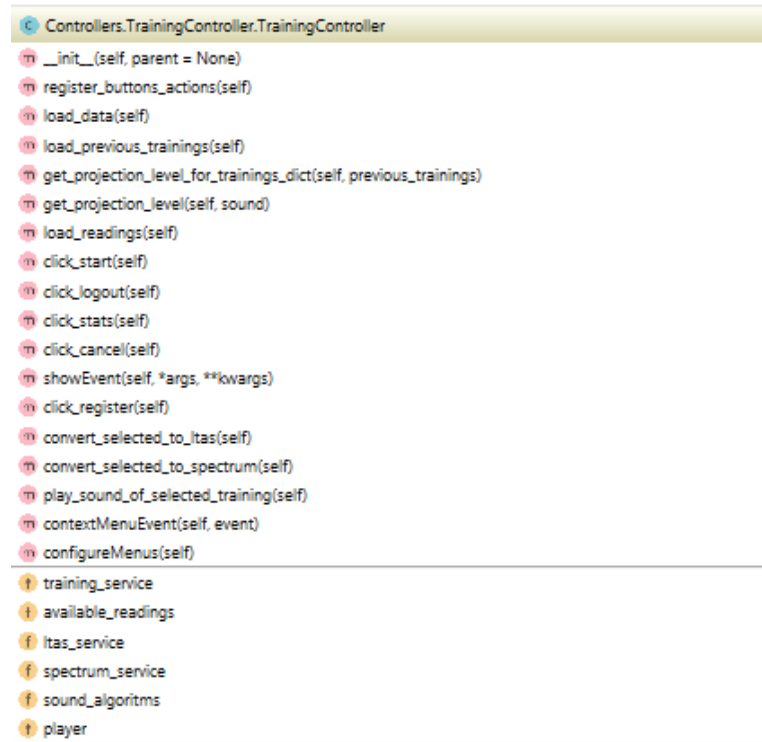


Figura 4.4: Diagrama de clases: Controlador trainings

- **Servicios:** Este módulo brinda servicios a los controladores. Es la parte más importante de la aplicación, ya que implementa el dominio de ella, haciendo uso intensivo de los algoritmos y repositorios sobre las entidades que se encuentra en el módulo de entidades. Aquí se definen los servicio de LTAS, Espectro, Entrenamiento y Usuario. También hay una clase Factory llamada `ServiceProvider`, encargada de crear y brindar un servicio con todos sus repositorios y algoritmos.

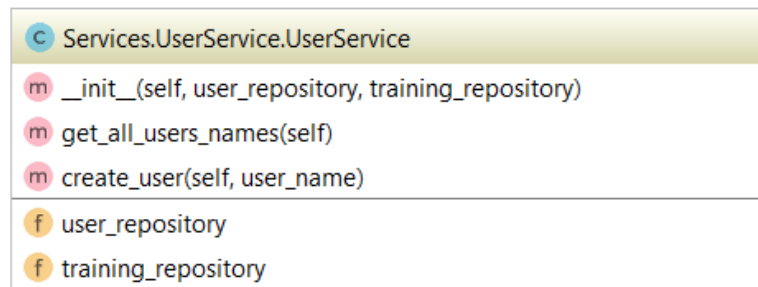


Figura 4.5: Diagrama de clases: Servicio de usuario

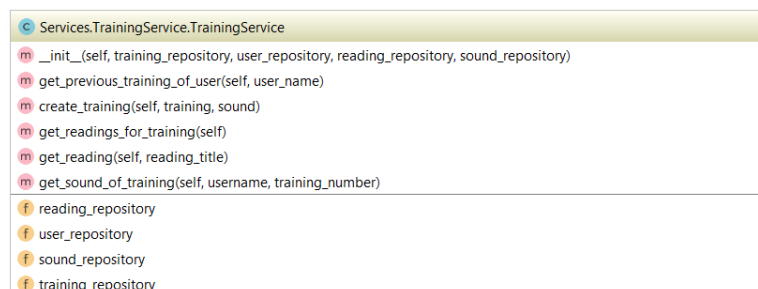


Figura 4.6: Diagrama de clases: Servicio de Entrenamiento

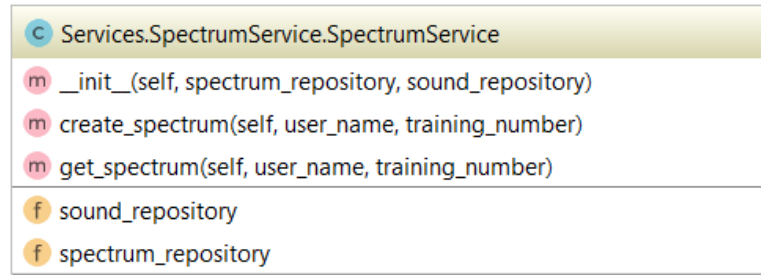


Figura 4.7: Diagrama de clases: Servicios de Espectro

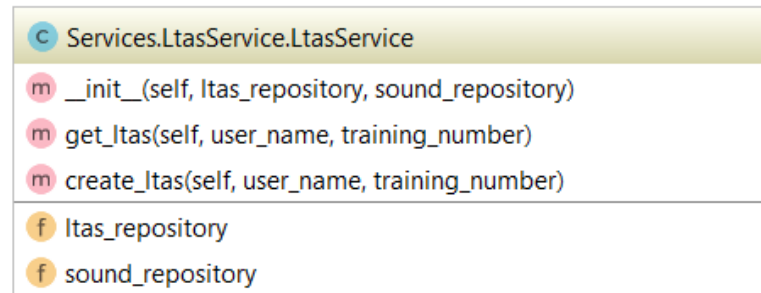


Figura 4.8: Diagrama de clases: Servicios de LTAS

- **Entidades:** Aquí se agrupan todas las entidades del sistema. Las entidades cuyos diagramas mostramos más abajo, son las encargadas de validar sus propios datos y representan los siguientes conceptos de la vida real en la aplicación: Espectro, Entrenamiento, LTAS, Usuario, Lectura, Audio.

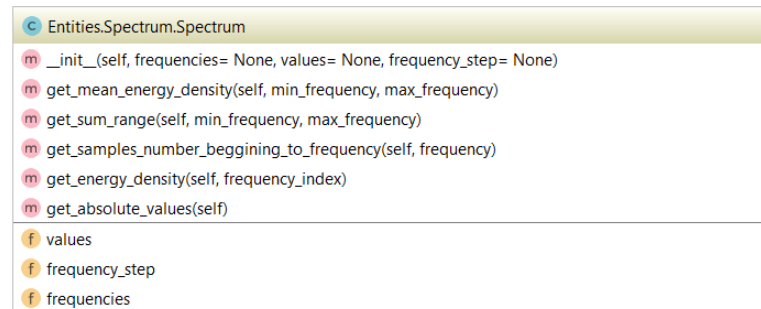


Figura 4.9: Diagrama de clases: Entidad Espectro

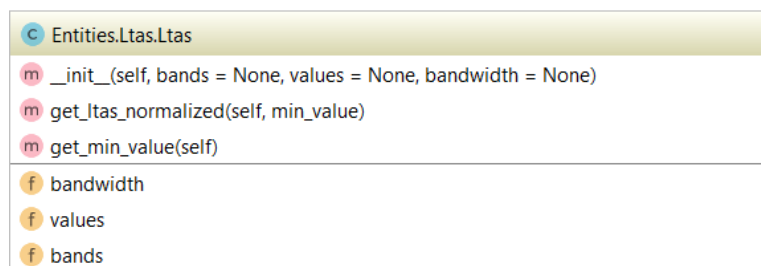


Figura 4.10: Diagrama de clases: Entidad LTAS

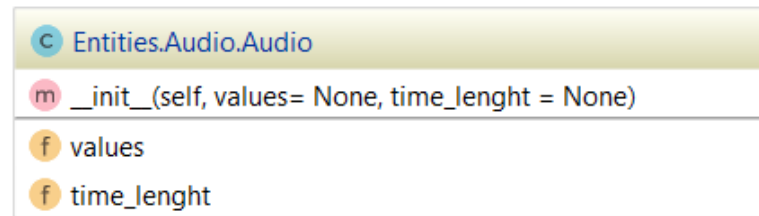


Figura 4.11: Diagrama de clase: Entidad Audio

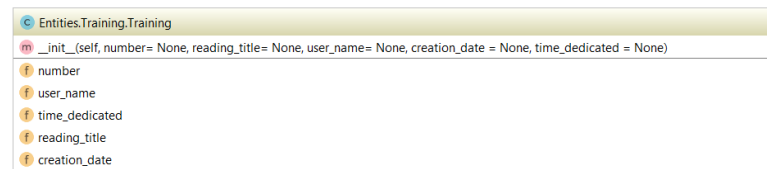


Figura 4.12: Diagrama de clases. Entidad Entrenamiento

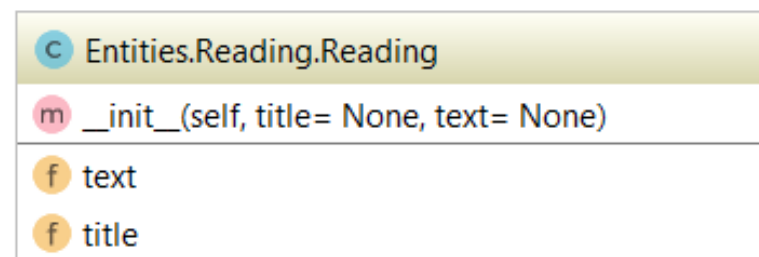


Figura 4.13: Diagrama de clases: Entidad Lectura

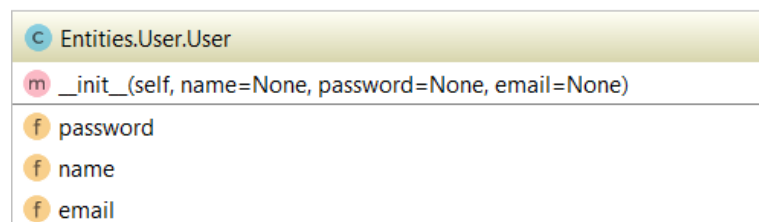


Figura 4.14: Diagrama de clases: Entidad Usuario

- Algoritmos:** El módulo de algoritmos como bien indica su nombre contiene los ficheros cuya responsabilidad son implementar los algoritmos de la aplicación, ya sea integrando la transformada de Furier de Numpy, calculando el espectro, el Ltas, o relacionando todo lo anterior para calcular la proyección. En el paquete de algoritmos usados por la aplicación están las clases creadoras de LTAS y Espectro, existiendo una clase destinada a los algoritmos comunes de sonido, y por último el Calculador de Proyección, que usa las clases anteriormente citadas para calcular la proyección.

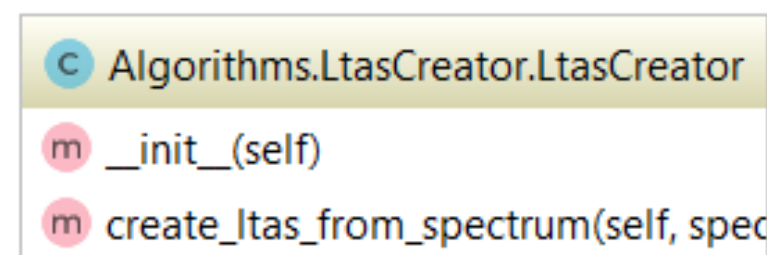


Figura 4.15: Diagrama de clases: Algoritmo Creador de LTAS

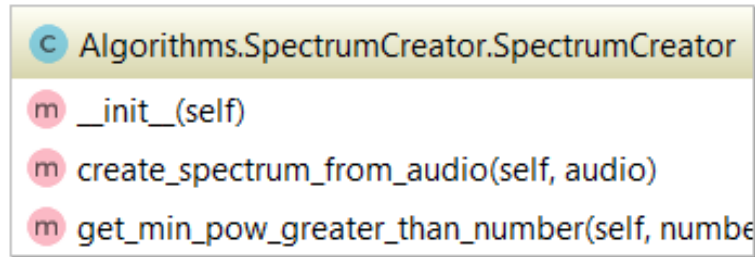


Figura 4.16: Diagrama de clases: Algoritmo Creador de Espectro

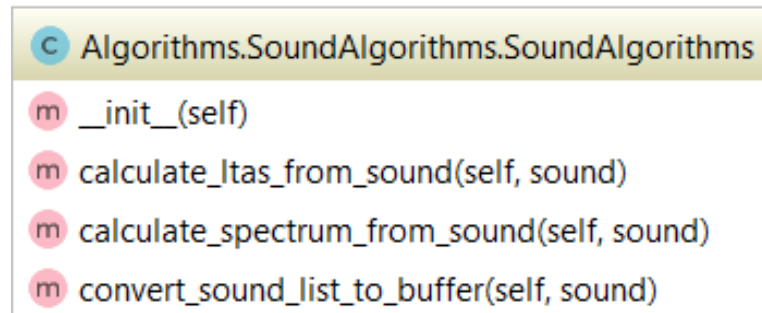


Figura 4.17: Diagrama de clases: Algoritmo de sonido

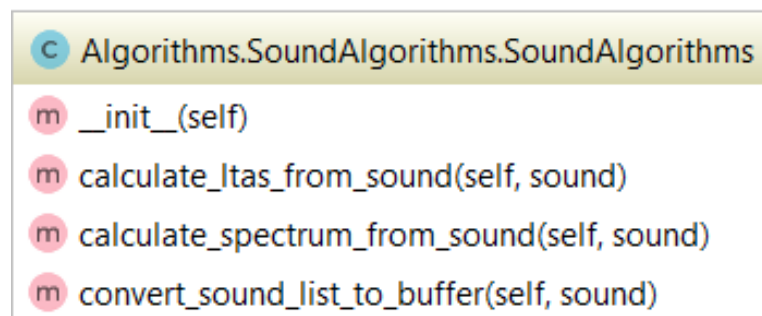


Figura 4.18: Diagrama de clases: Algoritmo calculador de proyección

- Repositorios:** Esto no es más que una abstracción de la capa de los datos, permitiendo cambiar el formato de los datos a una base de datos sql de cualquier tecnología o a una base de datos orientada a objetos. Solo habría que sustituir esta capa y ninguna otra parte de la aplicación se vería afectada. Este conjunto de clases implementan el acceso a datos para abstraer a los servicios de esta funcionalidad. Para cada entidad hay un repositorio que se encarga de las operaciones CRUD de esta. A continuación, mostramos dos ejemplos de repositorios, ya que en todo se repite el mismo esquema.

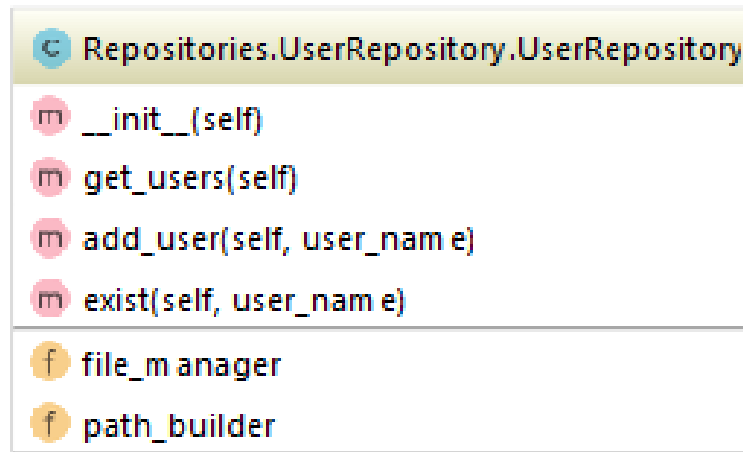


Figura 4.19: Diagrama de clases: Repositorio de Usuario

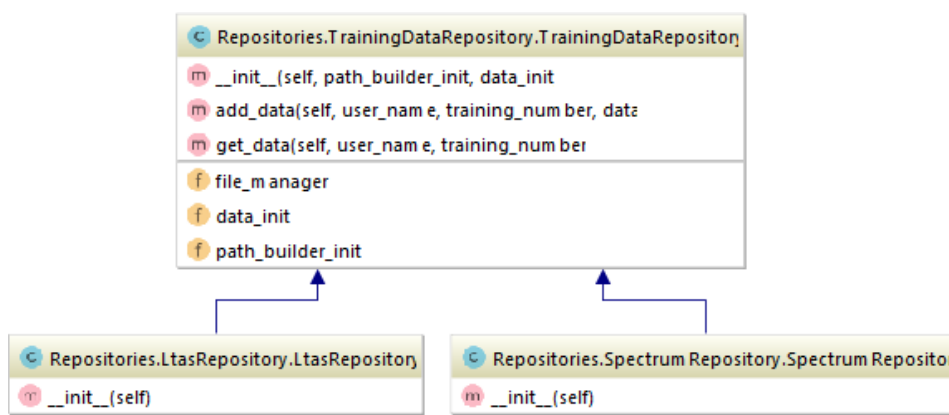


Figura 4.20: Diagrama de clases: Repositorios de Ltas y Espectro

Hay otras clases que han intervenido en la aplicación, pero no entraremos en detalle en ellas en el presente trabajo: Manejador de WAV, Reproductor de Sonidos, Grabador de Sonidos, Navegador de vistas, Proveedor de servicios, entre otras.

4.4. Diagrama de flujo de la aplicación

Al iniciar la aplicación el usuario puede registrarse él mismo o registrar a otros usuarios. Luego de hacer login, se procede a la ventana de entrenamientos, donde puede observar los datos de sus entrenamientos y también escuchar el sonido de cualquiera de ellos previamente, con solo seleccionarlo y elegir la opción: Play sound.

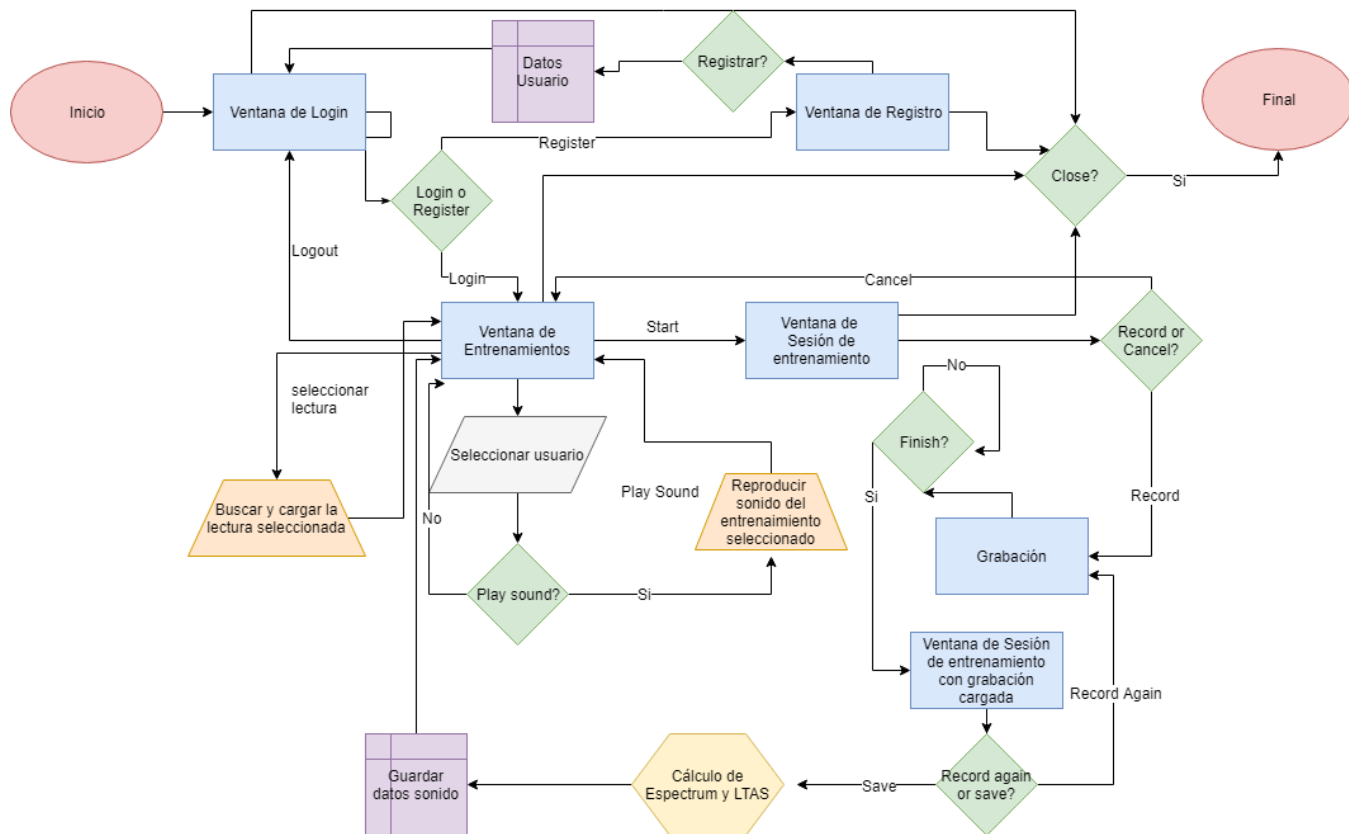


Figura 4.21: Diagrama de flujo de la aplicación

4.5. Descripción de la aplicación desarrollada

En este apartado describiremos con pantallazo de la aplicación la funcionalidad de la aplicación.

El usuario una vez que se registre y abra sesión (Login) en la aplicación, podrá ver en todo momento su histórico de entrenamientos con las proyecciones en una tabla y en gráfica. En la parte superior de la ventana, se encuentra lo necesario para comenzar a entrenar

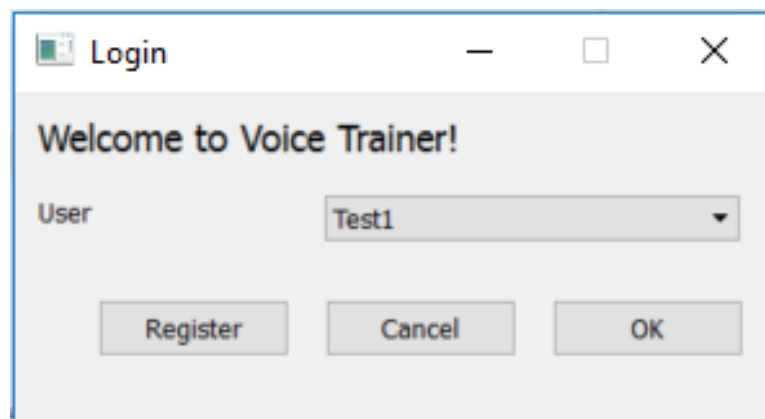


Figura 4.22: Aplicación: Ventana Login

En la siguiente imagen se puede ver que al seleccionar un entrenamiento de la lista, se muestran a la derecha las gráficas de su Espectro de Frecuencias y su LTAS.

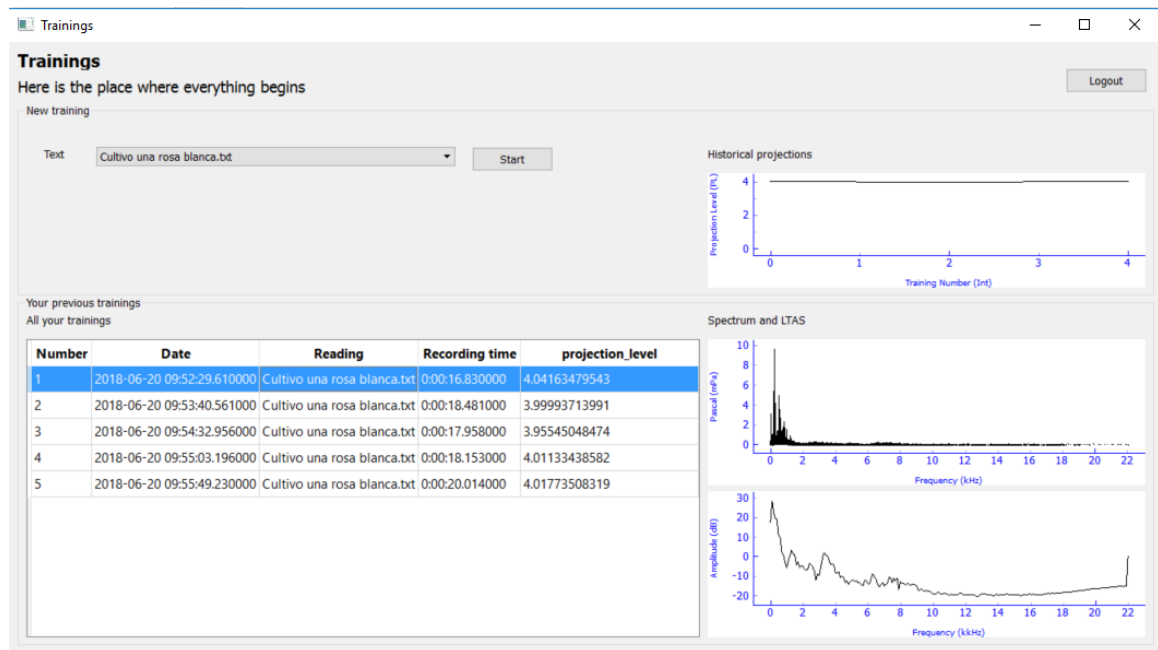


Figura 4.23: Aplicación: Ventana Entrenamientos Previos

Una vez que el usuario ha seleccionado un texto para leer, puede comenzar a entrenar.

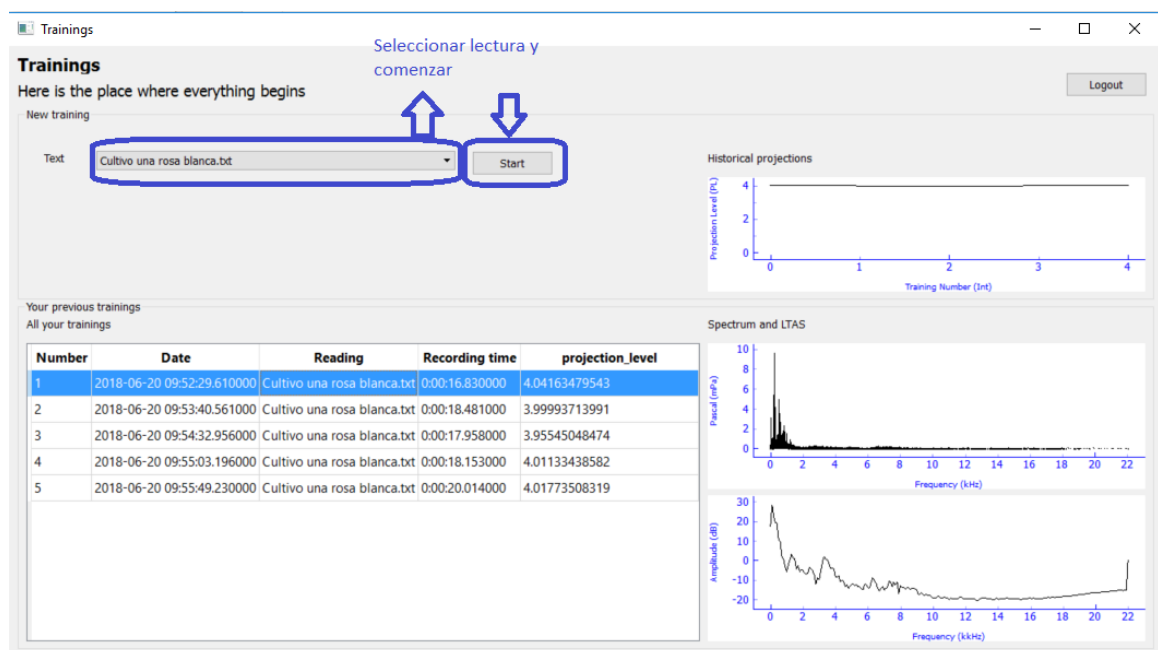


Figura 4.24: Aplicación: Comenzar nuevo entrenamiento

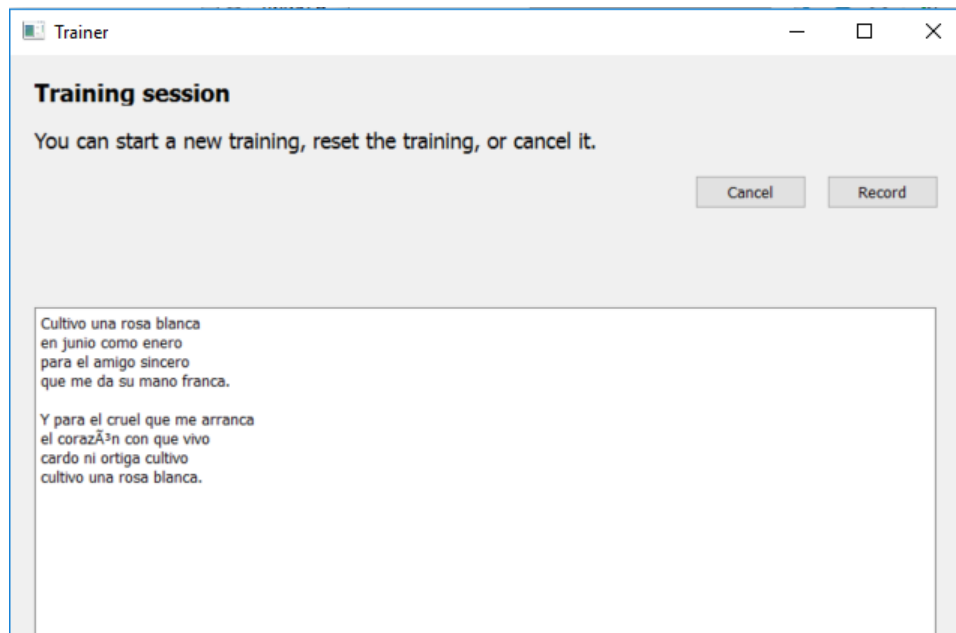


Figura 4.25: Aplicación: Ventana Sesión de Entrenamiento

El usuario tiene la posibilidad de repetir múltiples veces su entrenamiento, el cual una vez terminado se puede guardar. Poder repetir el entrenamiento es útil ya que podría ocurrir una interrupción o ruido ajeno a su voluntad, o incluso simplemente si desea modificar o mejorar posiciones del tracto bocal que le ayudaron en el pasado a proyectar mejor.

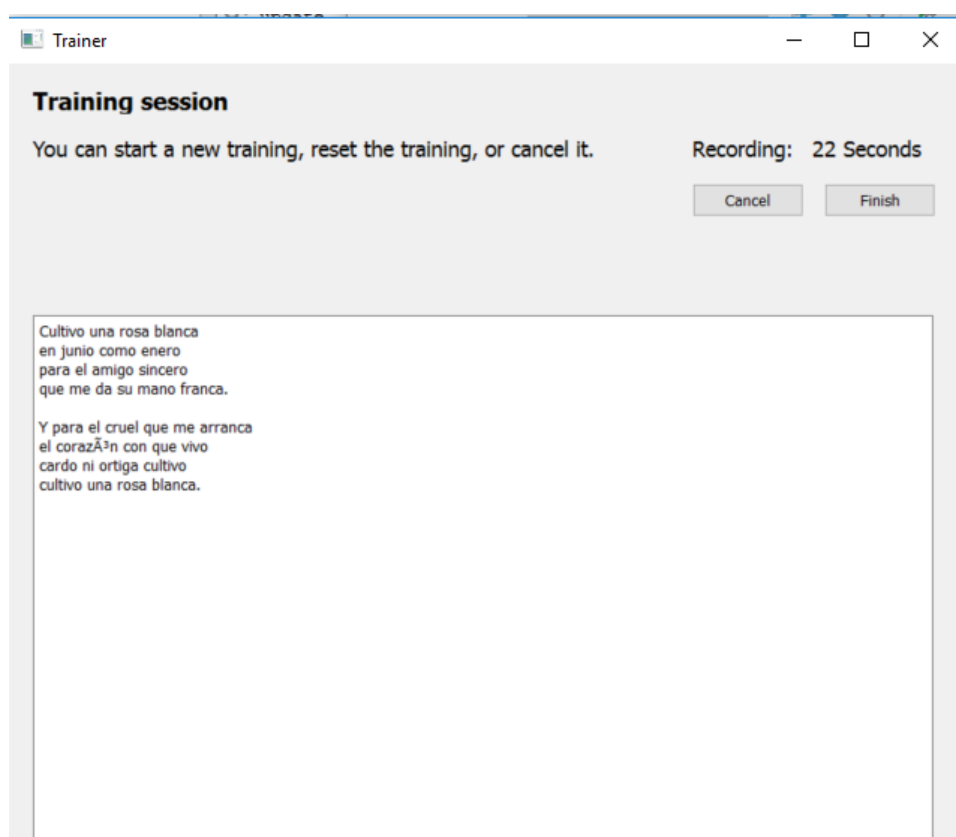


Figura 4.26: Aplicación: Grabando la lectura

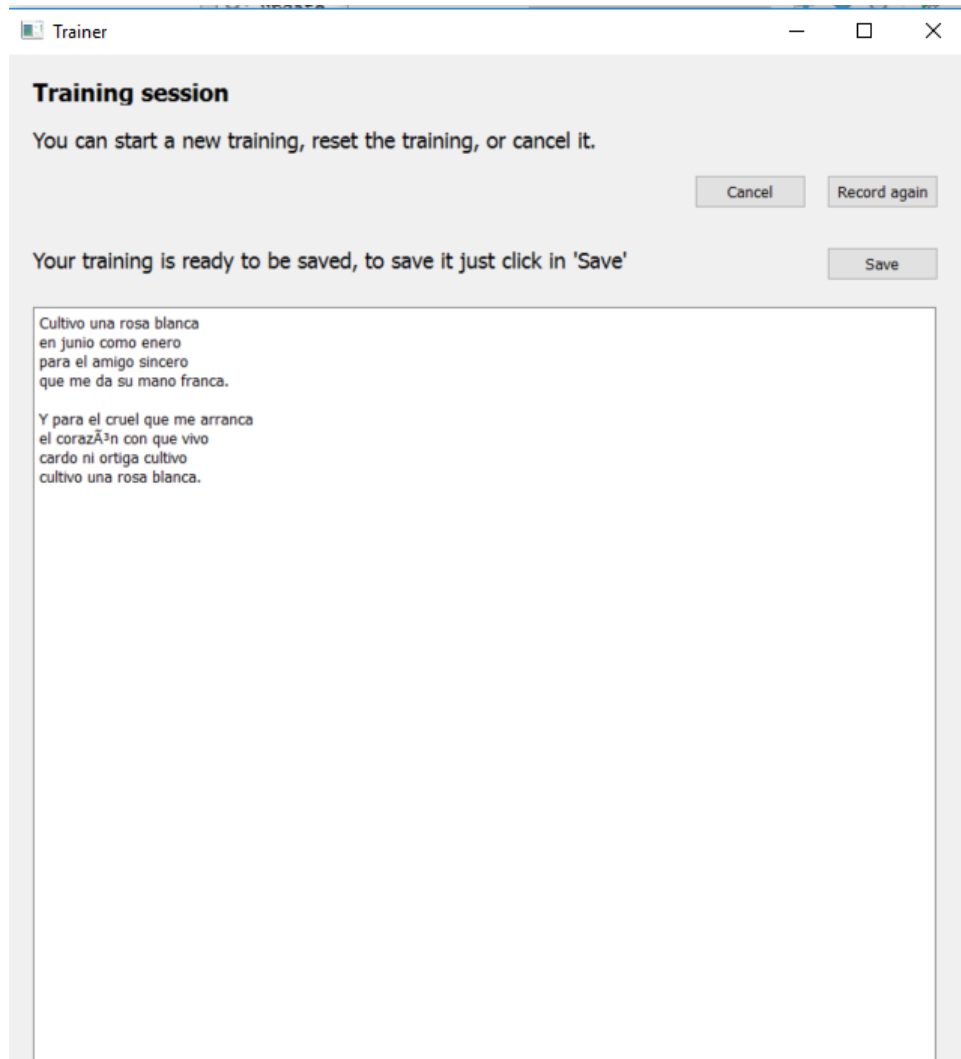


Figura 4.27: Aplicación: Grabación terminada

Por último al finalizar la sesión de entrenamiento, la aplicación le envía de vuelta a la ventana de entrenamientos, con las gráficas actualizadas.

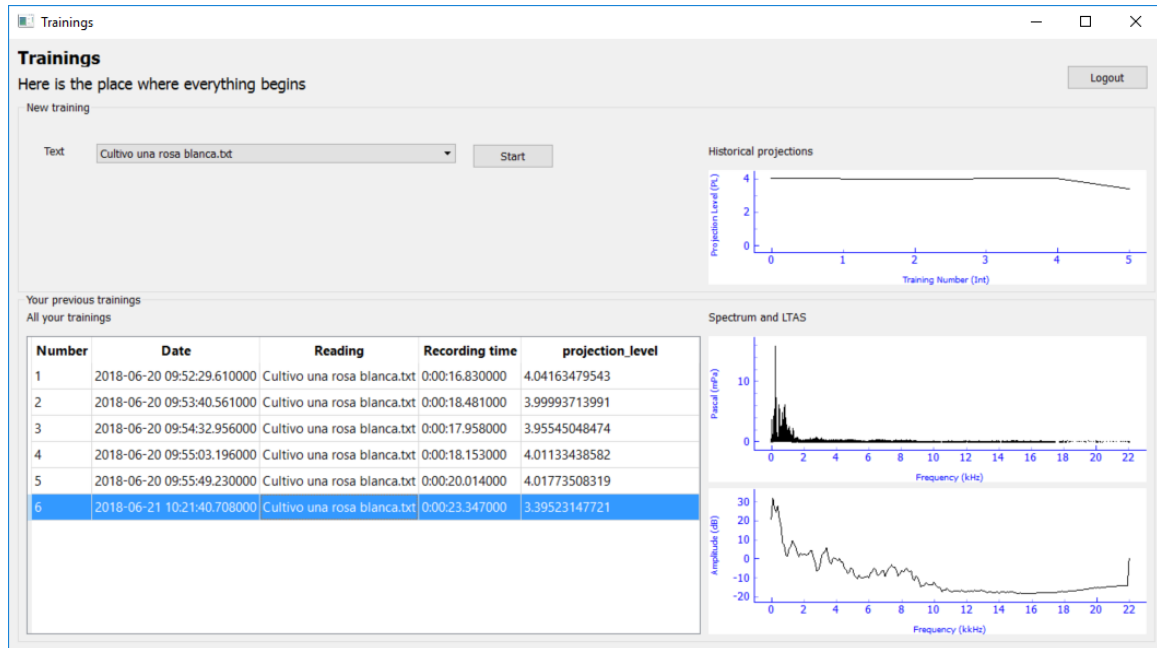


Figura 4.28: Aplicación: Entrenamiento añadido

4.6. Pruebas y Resultados

4.6.1. Test de los algoritmos Espectro y LTAS

Durante todo el proyecto se realizaron diversos tipos de tests. Comenzamos testeando la integración con Numpy y luego fuimos cubriendo los algoritmos a medida que los íbamos implementando.

Para la comprobación de los resultados usamos un programa utilizado por la comunidad de investigadores en estudios de fonética llamado PRAAT. Este programa está disponible de forma gratuita y mediante él se puede calcular el Espectro de un fichero de sonido y su LTAS.

Durante el desarrollo de los algoritmos de Espectro y LTAS creamos y ejecutamos varios tests. Debido a la gran cantidad de datos que manejan un Espectro y un LTAS, decidimos comparar estos con los resultados de una aplicación muy conocida en muchos de los estudios revisados: PRAAT. El PRAAT tiene su propia implementación del Espectro y Ltas, no pudimos usar dichas funciones pero si nos valió para comprobar la correctitud de nuestros resultados.

Usamos un test de valores específicos que luego eran comparados manualmente. Una grabación de 5 segundos tiene cientos de miles de valores, por tanto el tiempo empleado para verificar manualmente los resultados era muy alto, así que decidimos hacer dos tests que nos graficaran los valores y así ver de manera rápida si las gráficas coincidían con las gráficas del PRAAT, para después comenzar a hacer las comprobaciones valor a valor del fichero generado por el test y el fichero generado por PRAAT. Esta comparación debía ser manual, debido a pequeñas imprecisiones en algunos decimales.

Resumen de Tests para los algoritmos Espectro y LTAS:

1. Dos tests para graficar el Espectro y LTAS y comparar luego si coincidían visualmente con las gráficas del Praat.

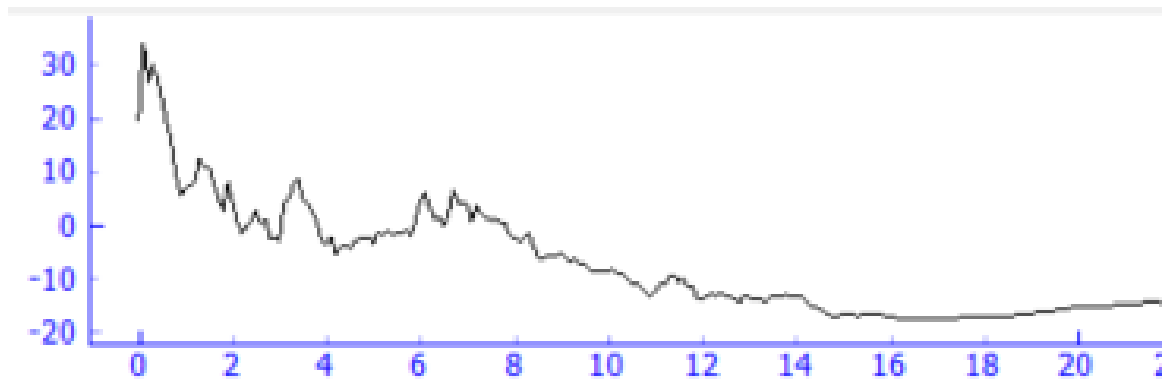


Figura 4.29: Comparativa LTAS: nuestra

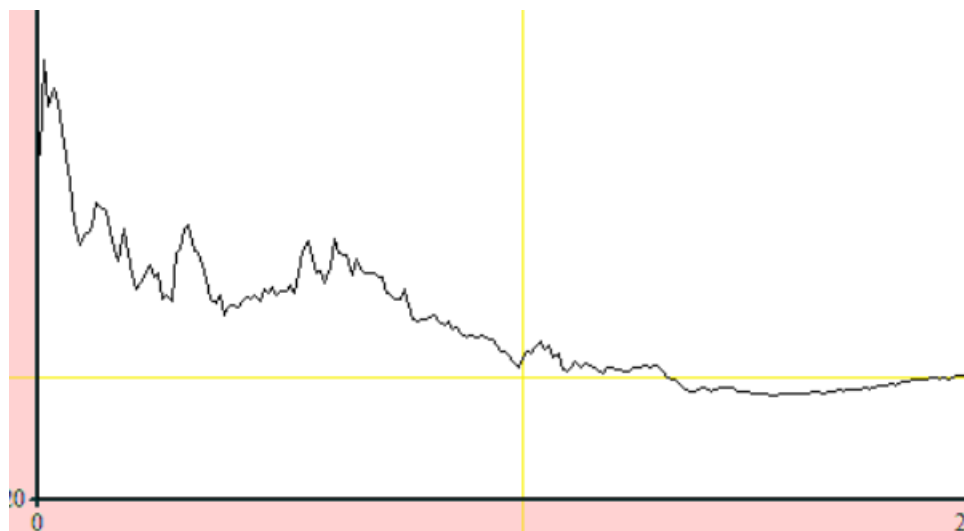


Figura 4.30: Comparativa LTAS: del PRAAT

2. Un test genérico que recibe el path de un fichero con los datos de entrada y el path de un fichero de salida. Luego estos valores se comparan con los resultados del praat.

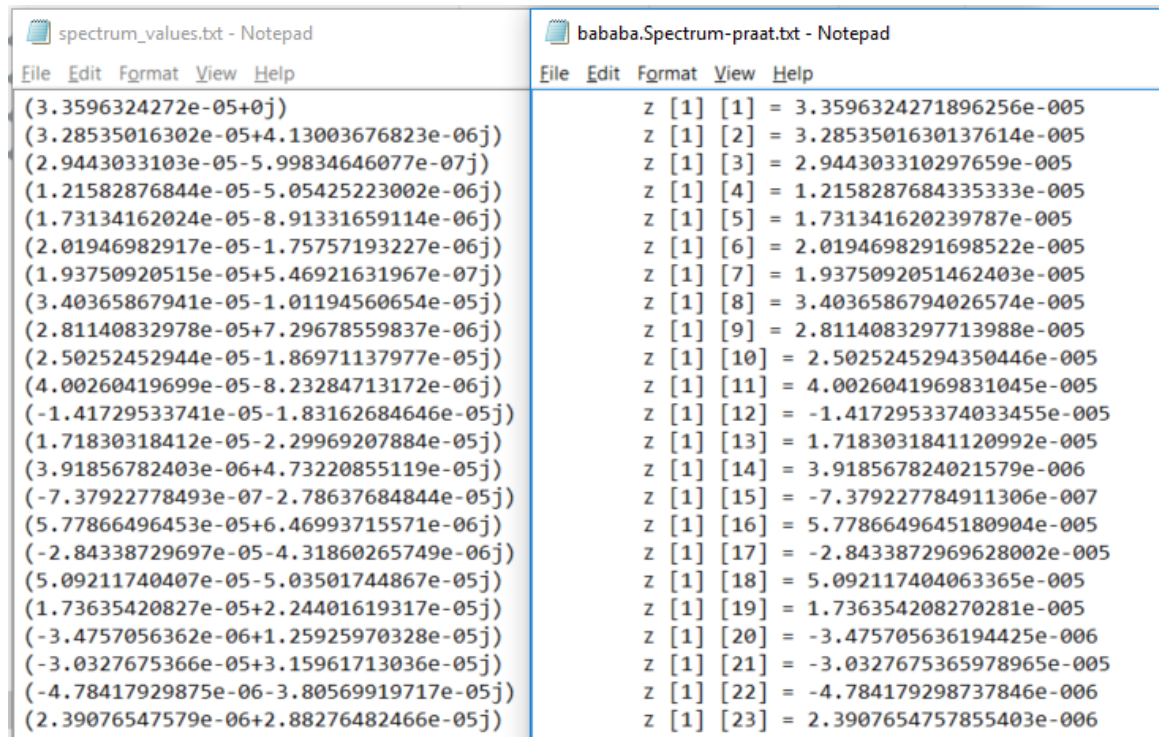


Figura 4.31: Comparativa de Espectro de Frecuencias

4.6.2. Test finales de los algoritmos y la proyección

Para comprobar una vez más la fiabilidad de los algoritmos y la implementación de la medición de proyección, realizamos diversos tests que detallamos a continuación.

1. Con un mismo usuario realizamos 5 entrenamientos usando la misma lectura, tratando de mantener el mismo tono y volumen de la voz y luego comprobamos que las gráficas y los valores de proyección eran similares. Valores de los test tratando de mantener la misma proyección:

5 Entrenamientos con igual proyección			
No. Entrenamiento	Lectura	Tiempo de grabación	Nivel de proyección
1	Cultivo una rosa blanca	16.83 s	4.041
2	Cultivo una rosa blanca	18.48 s	3.999
3	Cultivo una rosa blanca	17.95 s	3.955
4	Cultivo una rosa blanca	18.15 s	4.011
5	Cultivo una rosa blanca	20.01 s	4.017

Cuadro 4.1: Resultados de 5 entrenamientos con igual proyección

LTAS de los tests



Figura 4.32: LTAS del test 1

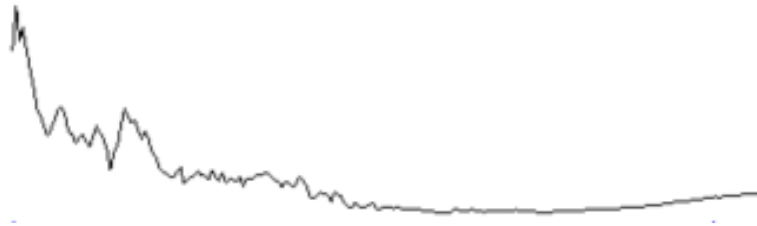


Figura 4.33: LTAS del test 2

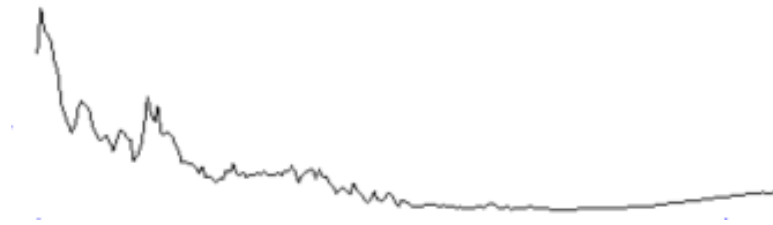


Figura 4.34: LTAS del test 3

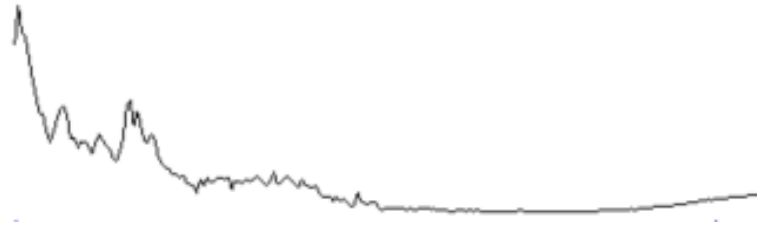


Figura 4.35: LTAS del test 4

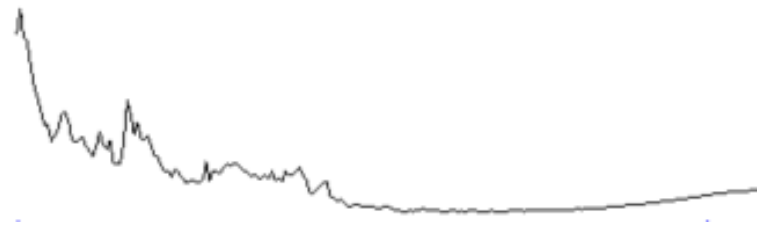


Figura 4.36: LTAS del test 5

2. Con varios usuarios realizamos un entrenamiento tratando de usar una voz apagada y luego otro entrenamiento intentando proyectar la voz, con el objetivo de verificar que los resultados de la proyección de la segunda con respecto a la primera eran mejores.

Test proyectando y no proyectando			
Intención de Proyección	Lectura	Tiempo de grabación	Nivel de proyección
Buena	Salmo 23	48.53 s	3.20
Mala	Salmo 23	41.94 s	3.00

Cuadro 4.2: Resultados de 2 tests: Proyectando y no proyectando

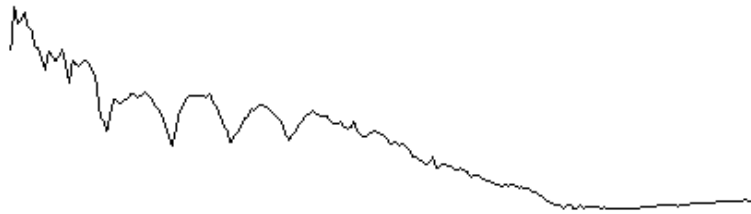
LTAS de los tests

Figura 4.37: LTAS del test proyectando

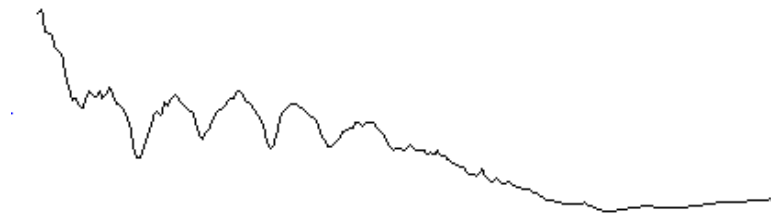


Figura 4.38: LTAS del test no proyectando

4.7. Encuesta de satisfacción

Realizamos una encuesta de satisfacción acerca de la aplicación la cual describiremos en este apartado.

4.7.1. Preguntas de la encuesta

- Valore las siguientes afirmaciones con valores entre 1 (nada de acuerdo) y 5 (totalmente de acuerdo):
 1. El diseño de la aplicación tiene un aspecto atractivo
 2. La aplicación es fácil de usar
 3. Las instrucciones proporcionadas son claras y se entienden con facilidad
 4. La aplicación proporciona información útil para saber si estoy proyectando la voz
- A continuación, escriba las sugerencias de mejoras que estime conveniente

4.7.2. Participantes de la encuesta

En esta encuesta participaron un total de 5 personas de diversas profesiones:

- Raquel García: Directora de un grupo vocal y con graves afecciones de la voz.

- Atanasio Mata: Licenciado en Sociología y conferencista de la Academia de Policía de Madrid.
- Adoniram Pires: Pastor y predicador de una iglesia.
- María Elena de la Moneda: Profesora Universitaria.
- Scot Musser: Director y profesor de niños 2-7 años en Colegio bilingüe.

4.7.3. Resultados de la encuesta

Media de valoración a las afirmaciones encuestadas

1. El diseño de la aplicación tiene un aspecto atractivo
Media: 4
2. La aplicación es fácil de usar
Media:4,6
3. Las instrucciones proporcionadas son claras y se entienden con facilidad
Media:4,6
4. La aplicación proporciona información útil para saber si estoy proyectando la voz
Media:4,25

Número de usuarios que valoran positivamente cada afirmación

Definimos por una afirmación con un valoración positiva, aquellas afirmaciones donde el usuario valore con 4 o 5. Abajo mostramos la cantidad de usuarios por afirmación que han valorado positivamente.

1. El diseño de la aplicación tiene un aspecto atractivo
No. Usuarios:4
2. La aplicación es fácil de usar
No. Usuarios:5
3. Las instrucciones proporcionadas son claras y se entienden con facilidad
No. Usuarios:5
4. La aplicación proporciona información útil para saber si estoy proyectando la voz
No. Usuarios:4

Recomendaciones de los usuarios a mejorar en la aplicación:

- Sería interesante poder comparar mi voz con la de alguien que proyecte muy bien, de manera que acelere el proceso de entrenamiento.
- Me parece muy claro y una herramienta muy buena para el desarrollo de la voz
- Para personas con problemas de visión es difícil leer la letra
- Me parece confusa la curva del histórico de proyecciones

5

Conclusiones y trabajo futuro

5.1. Conclusiones

El propósito de este Trabajo de Fin de Grado como ya se ha expuesto desde el inicio, es proveer una herramienta que ayude al usuario a mejorar la proyección de la voz.

Para lograr este objetivo revisamos y analizamos múltiples trabajos e investigaciones de diferentes autores que trabajan este campo, para una vez comprobada que la proyección de la voz era real, intentar desarrollar la aplicación. Los requisitos iniciales fueron cumplidos en su creación. Se ha utilizado Python, Numpy, PyAudio, Git(con Github), todos ellos software libre, con lo cual si alguien quiere más adelante sumarse al trabajo y extenderlo o mejorarlo puede hacerlo, teniendo acceso gratuito mediante Github.

Como mostramos, se realizó una encuesta de satisfacción a un pequeño grupo; fueron escogidas 5 personas que tienen relación con el tema que trabajamos “la proyección de la voz”, y tomamos notas a seguir de sus sugerencias.

El desarrollo de este trabajo ha reforzado en el autor del mismo, conocimientos dados en la carrera como por ejemplo la Transformada de Fourier (reconociendo su utilidad en aplicaciones prácticas), Programación Orientada a Objetos(Python), Ingeniería de Software, habilidades investigativas, etc.

La parte investigativa inicial del proyecto supuso un esfuerzo considerable y el número de horas dedicado a dicha investigación ha sido también encomiable. En principio no sabíamos siquiera si la proyección era posible y si se podía medir matemáticamente y aunque intuitivamente era algo que suponíamos, no dejamos de tener cierto grado de incertidumbre al respecto, por lo que tuvimos que buscar, leer, traducir, extraer información válida y también descartar numerosos estudios de la proyección de la voz que no fueron utilidad, tal y cómo se muestra en el apartado de Referencias Bibliográficas.

5.2. Trabajo Futuro

El trabajo futuro de esta aplicación estará determinado por su impacto. Consideramos conveniente seguir dedicándole tiempo a esta aplicación, en principio para darle mantenimiento. También sería apropiado incorporar expertos en el área y seguir perfeccionándola, quizás como

parte de un máster o un doctorado, este al menos es el sueño del autor.

Por otra parte, continuaremos haciendo cambios y mejoras, teniendo en cuenta inicialmente los resultados de las encuestas, en que por ejemplo se sugirió que la interfaz gráfica sea más agradable a la vista. Otro aspecto a mejorar (a criterio del autor) sería hacer aún más precisa la medición de la proyección, pero esto dependería de que se publiquen estudios que confirmen el esfuerzo vocal sobre 5 kHz- 8 kHz. Incluir ejercicios que recomendarle al usuario en la aplicación y buscar voces con una proyección bien definida que le sirva para guiarse es otra ventaja en la que pudiéramos trabajar.

Otra manera de contribuir al desarrollo de la aplicación sería editando la voz del usuario para reforzarle la zona de proyección 3 kHz- 4 kHz, así podría imitar su propia voz, algo mucho más natural y lógico para el usuario que tratar de imitar la voz de otra persona.

En resumen, podemos inferir que tal y como continúan las investigaciones a nivel global acerca de la proyección de la voz, este trabajo podría tener numerosas mejoras. Hay que añadir que tanto el autor como el tutor, han disfrutado y aprendido muchísimo en el proceso.

Glosario de acrónimos

- **LTAS:** Long-term Average Spectrum

Bibliografía

- [1] Leyno T. Long-term average spectrum study on speaking voice quality in male voices. *SMAC93 Proceedings of the Stockholm Music Acoustics Conference July 28-August 1*, July 28-August 1 1993.
- [2] Jeans J.H. Science & music. *New York : Macmillan Co.*, page 104, 1938.
- [3] Titze I.R. Principles of voice production. *Englewood Cliffs, N.J: Prentice Hall*, 1994.
- [4] Wood S. What are the formants, beginners guide to praat, phonetics lab, lund university.
- [5] Johan Sundberg. The acoustics of the singing voice. *Scientific American*, page 82, Mar 1977.
- [6] Bele I.V. The speaker's formant. *Journal of Voice*, Vol. 20, No. 4, 2006.
- [7] Mario Fleischer, Silke Pinkert, Willy Mattheus, Alexander Mainka, and Dirk Mürbe. Formant frequencies and bandwidths of the vocal tract transfer function are affected by the mechanical impedance of the vocal tract wall. *Biomechanics and Modeling in Mechanobiology*, 14(4):719–733, Aug 2015.
- [8] R.P. Daga and A.M. Panditrao. Acoustical analysis of pain cries' in neonates: Fundamental frequency. *ICEICE No.3*, 2011.
- [9] Hannah Elisabeth Stoffel. Acoustic analysis of actresses' voices. *Universiteit Utrecht*, 2011.
- [10] Jonas Lindh. Visual acoustic vs. aural perceptual speaker identification in a closed set of disguised voices. *Proceedings, FONETIK*, 2005.
- [11] Tomi Kinnunen, Ville Hautamäki, and Pasi Fränti. On the use of long-term average spectrum in automatic speaker recognition. *International Symposium on Chinese Spoken Language Processing (ISCSLP 2006)*, December 13-16, 2006.
- [12] Boersman and Weensign. Praat: Doing phonetics by computer. *Retrieved from <http://www.praat.org>*, 2005.



ANEXO: Encuesta de la aplicación

ENCUESTA DE SATISFACCIÓN A USUARIOS

Acerca del uso de esta herramienta de entrenamiento vocal y con la finalidad de conocer su grado de satisfacción con la aplicación usada, le agradeceríamos respondiera con sinceridad las preguntas que a continuación aparecen.

- Valore las siguientes afirmaciones con valores entre 1 (nada de acuerdo) y 5 (totalmente de acuerdo):
 1. El diseño de la aplicación tiene un aspecto atractivo:
1_ 2_ 3_ 4_ 5_
 2. La aplicación es fácil de usar:
1_ 2_ 3_ 4_ 5_
 3. Las instrucciones proporcionadas son claras y se entienden con facilidad:
1_ 2_ 3_ 4_ 5_
 4. La aplicación proporciona información útil para saber si estoy proyectando la voz:
1_ 2_ 3_ 4_ 5_
- A continuación, escriba las sugerencias de mejoras que estime conveniente: